Best Available
Copy
for all Pictures

AD-777 993

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
Image Processing Institute, Electronic Sciences Laboratory, University of Southern California, University Park, Los Angeles, California 90007.	UNCLASSIFIED
	2b. GROUP

3. REPORT TITLE

THE ROLE OF PHASE IN ADAPTIVE IMAGE CODING

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report, December 1973

5. AUTHOR(S) *(First name, middle initial, last name)*

Andrew G. Tescher

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1973		49
8a. CONTRACT OR GRANT NO. F08606-72-C-0008	9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO. ARPA Order No. 1706	USCIPI Report 510	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	
d.	None	

10. DISTRIBUTION STATEMENT

Approved for release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

This technical report includes a detailed analysis of image modeling aspects of the transform coding problem. Two alternate prediction algorithms are analyzed for the transform sample variance estimation: the first technique uses a two-dimensional polynomial to model the image power spectral density; the second technique is a simple recursive approach based on previously quantized values.

The generalized phase concept is developed and plays a vital role in the coding algorithms. Both the Fourier and Walsh transforms are utilized, the former being demonstrated to have superior performance. A non-negative image constraint is explored via the Lukosz bound.

The experimental phase of the study includes two dimensional coding of monochrome, and three dimensional coding of color, as well as interframe images with coding at 0.38, 0.55, and 0.25 bits/pixel, respectively. It is demonstrated that adaptive transform domain modeling is important, and that large-size transforms, in conjunction with the proper image model, can significantly outperform block-encoding techniques.

DD FORM 1473
1 NOV 65

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Image Coding, Fourier Transforms, Walsh Transforms, Phase Coding, Generalyzed Phase, Frame to Frame Coding, Color Image Coding, Lukosz Bound; Polynomial Fit.						

THE ROLE OF PHASE IN ADAPTIVE
IMAGE CODING

Technical Report

by

Andrew G. Tescher

Image Processing Institute

Electronic Sciences Laboratory

University of Southern California
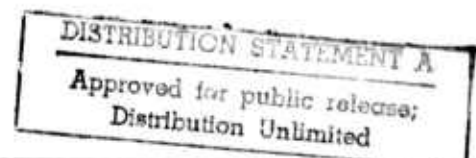
University Park

Los Angeles, California 90007

December 1973
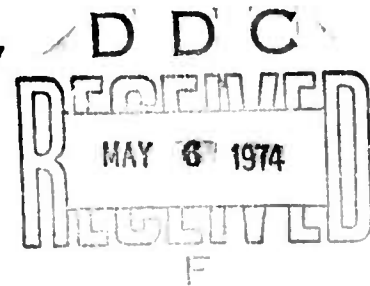
DDC

RECEIVED

MAY 6 1974

F

The views and conclusions in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

ABSTRACT

Results of a comprehensive research program to develop efficient transform image coding algorithms are reported in this dissertation. The objective is to develop algorithms that outperform the conventional block-encoding procedures, i.e., achieve data rates below the one bit/picture element which is the approximate lower limit for conventional transform coders.

The dissertation includes a detailed analysis of image modeling aspects of the transform coding problem. Two alternate prediction algorithms are analyzed for the transform sample variance estimation; the first technique uses a two-dimensional polynomial to model the image power spectral density; the second technique is a simple recursive approach based on previously quantized values. The actual coding algorithms utilize the latter approach.

The generalized phase concept is developed and plays a vital role in the coding algorithms. Both the Fourier and Walsh transforms are utilized, the former being demonstrated to have superior performance. A non-negative image constraint is explored via the Lukosz bound.

The experimental phase of the study includes two dimensional coding of monochrome, and three dimensional coding of color, as well as interframe images with coding at 0.38, 0.55, and 0.25 bits per pixel, respectively. It is ascertained that decoded and reconstructed images are not significantly degraded. It is also demonstrated that adaptive transform domain modeling is important, and that

large-size transforms, in conjunction with the proper image model, can significantly outperform block-encoding techniques.

A requirement for large-size transforms can easily discourage hardwired usage. Techniques can be developed, however, that could advantageously be employed for computer-to-computer image transfer.

Although the new coding-decoding methods are sensitive to channel errors, it is demonstrated that they produce data which are statistically equivalent to a discrete memoryless source. Thus, conventional channel coding techniques can be used.

TABLE OF CONTENTS

LIST OF FIGURES

viii

LIST OF TABLES

ACKNOWLEDGEMENT

xi

I. INTRODUCTION

The human visual system can absorb and evaluate vast amounts of pictorial information. The range of the visual data includes many different classes such as graphics, biomedical images or aerial photographs. The human eye responds to color as well as intensity; consequently the general description of an image also contains spectral information. If the time history of the image is to be characterized, the dimensionality of the description is further increased.

Mathematically, an image can be represented by a function of four variables, $I = I(x, y, t, \lambda)$. The spatial coordinates are x, y, the variable t represents time and λ is the wavelength representing a particular spectral component. The I represents the energy to which the eye as a photoelectric detector responds. The energy is a non-negative quantity; consequently, the following constraint must be satisfied for an image

$$I(x, y, t, \lambda) \geq 0$$

This simple non-negative constraint introduces various additional constraints for image sampling and filtering.

This dissertation is devoted to an adaptive technique of image coding. In terms of the definition of an image, image coding is specified as a process by which the <u>analog</u> image function I is represented as a sequence of binary digits. Clearly, the binary representation must be unique and invertable for a given coder. The relative

1

efficiency of image coders can be directly compared in terms of the binary digit sequence length generated to characterize a given image.

1.1 Review of Coding Objectives, Techniques, Results

Although the primary objective of image coding has been communication bandwidth reduction for pictorial data, there are additional equally important considerations. The general availability of increasingly powerful digital computers has permitted numerical implementations of many image operations. The degrees of freedom in a typical image are quite large; consequently, the storage and access of pictorial data itself represents a significant problem.

The definition of image coding given on page 1 is essentially a source coding process. A schematic of the simplified communication system is given in Figure 1.1-1.

It is the source encoding/decoding which is relevant to the nature of pictorial information. Specifically, an efficient source coding process will utilize the statistics and dimensionality (space, time, and color, as previously indicated) of the pictorial data. The conversion of the analog image into a binary stream involves various distinct steps which may include an analog one- or two-dimensional prefilter, sampler, quantizer, digital preprocessor, and statistical encoder. All of these operations are largely determined by the nature of the source.

The channel encoding/decoding, unlike source encoding/decoding, should be insensitive to the original character of the data.
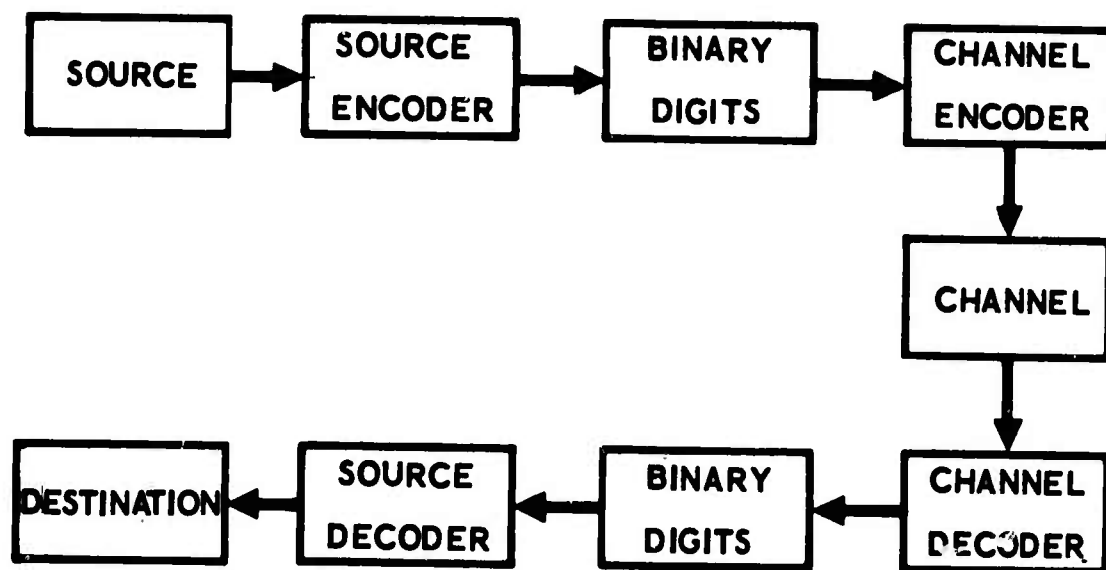
Figure 1. 1-1.   Classical Communication Model

Although consideration of the channel and related parameters is important in the overall communication problem, the relevant encoding/ decoding process is not unique to the nature of pictorial data. It can be expected that the source encoding process will produce a sequence of binary digits which are statistically equivalent to a set produced by a memoryless discrete source. Classical channel encoding techniques should, therefore, be applicable to the source coded image data without specific reference to the pictorial nature of the information.

Although the sampling and quantizing process should be considered as an integral part of picture coding, it is rarely done. Conventionally, the input to the coder is a sampled and quantized image which the coding algorithm will process such that its output consists of a reduced number of binary digits. The conventional sampling is performed over a rectangular grid and the analog samples are quantified to 64-256 quantum levels. A picture coding algorithm reduces the source rate, or equivalently the transmission bandwidth requirement, by reducing the number of samples and/or reducing the number of quantum levels.

The well-known and accepted technique of differential pulse code modulation (DPCM) reduces the number of quantum levels without sample reduction (Cutler, 1952; Graham, 1958; O'Neal, 1966). DPCM achieves the rate reduction by encoding sample differences rather than the samples themselves. Many different categories exist for this coding technique. Compared with the 8-bit conventional

PCM code, a well-designed DPCM system can achieve a factor of three rate reduction, 2.5-3 bit per original picture element.

The various algorithms that decompose the image or its derivatives into contours can achieve significant rate reduction for specific types of images, namely, the ones that can be described by a few number of contours. The disadvantage of this technique is the high degree of computational complexity and large buffer requirement (Graham, 1967). This requirement is that the entire image must be simultaneously available to the processing algorithm. Contour tracing algorithms have been adapted to frame-to-frame image coding (Habibi, 1973). In this case, the frame-to-frame image difference is subjected to the coding algorithm. The receiver, upon decoding the difference image, updates the previous frame. Frame-to-frame coders of this type can achieve a rate of one bit.

Coders that adapt to the local statistics of the image can achieve additional rate reduction over nonadaptive algorithms. The dual coder is an example of this technique (Frei, Schindler, and Veitinger, 1972). In this case, the sampling rate is changed according to the amount of local picture detail.

As stated earlier, the general image representation requires four dimensions, two for space, one for time, and one for color. Most coding techniques consider only monochrome images. Only recently has color coding acquired more attention (Bhushan, 1970; Pratt, 1971). Use of frame-to-frame redundancy in images is another research topic which has not been extensively explored.

1.2 Transform Techniques for Image Coding

Most classical spatial domain coding techniques (contour coding is the exception) generate code words based on the original picture elements (PEL) through a one-to-one mapping. In other words, the bandwidth reduction is achieved by requantization. Although the mapping is one-to-one, inter-element correlations are often utilized by the coding algorithm (Habibi, 1971). What is fundamentally different for transform coding is that part or all of the image is transformed into another domain via an invertable mapping. The sample reduction and requantization are performed on the transformed values and the resultant code words are then transmitted through the channel. The receiver will attempt to reconstruct the original image utilizing the inverse of the transform upon receiving the appropriate code words.

Numerous techniques have been developed for transform coding over the last five years (Wintz, 1972). Although practical ranking cannot be made, many of these techniques result in data rates as low as 1 bit/pel. The theoretical justification and motivation behind transform coding has been rather varied. Transform coding has been analyzed essentially by statistical tools. One basic motivation has been sample reduction. The "useful" transforms have the property that most of the image energy is concentrated in relatively few transform samples. Stating it differently, many transform samples have very small amplitudes and can therefore be discarded without being transmitted through the channel.

The Karhunen-Loeve (K-L) transform is the optimum transform for images describable by second order statistics (Thomas, 1968). It has been shown that for correlated Gaussian sources the optimum quantizer will uncorrelate the samples via the K-L decomposition and the bit rate is determined in proportion to the transformed variance samples. The K-L transform, by definition, diagonalizes the image covariance function. The diagonal terms are the eigenvalues and are ordered in decreasing magnitude.

The K-L transform is almost synonymous with optimum image coding, and sometimes the relevant assumptions are neglected. In the practical sense, K-L transform has somewhat less universal utility. Even theoretically, the K-L transform is optimum in the mean-square error sense and only through second order statistics. For a correlated Gaussian source, the optimality is achieved in fact. Practical image sources are not Gaussian and have higher than second order moments which cannot be derived from the first two.

The lack of availability of the covariance function is another difficulty. There are two fundamental questions to be analyzed: (1) How meaningful is the concept of covariance to images? Stated in another way: is image covariance a valid statistical concept for images which are likely to be nonstationary? (2) If we ignore the first question, how will the functional form of the covariance function be determined?

Question number one is, in fact, ignored in practice; and the perhaps oversimplified statement can be offered that because of the

lack of better statistical understanding of images, no better param-
eter has yet been offered.

The approximation of the covariance function or its transform
domain equivalent can be done either numerically or by a closed
form function. For the first case, one can directly determine the
transform sample variances experimentally and make the bit-
assignment accordingly. An example for the functional form is:

$$\exp\left(-\alpha\left|x\right|-\beta\left|y\right|\right)$$

This simple experimental form has been used successfully in spite
of its gross simplicity (Habibi and Wintz, 1972). The parameters
α and β represent the horizontal and vertical correlation, and
directional separability of these principal axes is assumed. The
exponential form of the covariance function is attractive. It is
simple and the parameters α and β are easily estimated.

A small number of statistical parameters is desirable in any
coding scheme. Since both the receiver and transmitter must know
these parameters, their transmission may require non-negligible
bandwidth and should be considered as part of the overall bit rate.

The separable form of the covariance function, although not
necessarily characteristic of actual image fields themselves, has
served a useful purpose.

Let

$$R(x, y) = R_x(x)\, R_y(y)$$

be the covariance function. Let T represent the transformation operator, then, symbolically, if $T = T_x T_y$

$$T\{R(x, y)\} = T_x\{R_x(x)\} T_y\{R_y(y)\} = S_x(u)S_y(v)$$

where

$$T_x\{R_x(x)\} = S_x(u)$$

$$T_y\{R_y(y)\} = S_y(v)$$

The generalized power spectral densities S_x, S_y should decrease for increasing values of the transform domain coordinates, u, v if the transform operations are to be useful for image coding. This fact is achieved by the proper choice of the transform operator T. The bit assignment is proportional to $\log S_x(u) + \log S_y(v)$. The clear implication is that the principal axis in the transform plane (e.g., when either u or v is zero) will receive a relatively large fraction of the available bits. Most transform image coding techniques operate on adjacent sub-blocks rather than the whole image itself. The separable covariance function results in effective superior recovery of the horizontal and vertical image structure. The block boundaries, however, become an integral part of the image statistics and their objectionable visual appearance is greatly diminished by the utilization of the separable covariance model. On the other hand, at very low bit rates, excess amount of the bandwidth may be required to

maintain the horizontal and vertical structure at the expense of a greater amount of resolution loss than may be justified.

The choice of the actual transforms have been dictated by the requirement of computational ease and the potential of practical implementation. The transformations, considered to date, are Fourier, Hadamard, K-L, (on sub-blocks) and more recently, the Slant transform. (Habibi and Wintz, 1971; Anderson and Huang, 1971; Pratt, 1972; Pratt, Welch and Chen, 1972.) All of these except K-L can be implemented by "fast" algorithms.

1.3 Research Objectives

The amount of visual data generated in commercial and scientific applications is enormous. The ordinary home television set generates over 500 x 500 samples 30 times a second. The Earth Resources Technology Satellites and weather satellites typically produce in excess of 4000 x 4000 and 8000 x 8000 data points, respectively. Data storage and transmission becomes a major problem for pictorial sources because of the excessive amount of data. Clearly, techniques that permit greater efficiency (e.g., reduction in the required bandwidth) are urgently needed.

Numerous approaches have been considered for efficient picture coding. While these techniques are based on widely different considerations, they are all motivated by the required simplicity of potential implementation. Consequently, the developed algorithms are relatively simple, utilize simple models, and are somewhat inflexible in terms of their adaptivity to the image structure.

The philosophy on which this dissertation is based, emphasizes flexibility and maximum efficiency at the possible expense of increased computation and buffer requirements. Demonstration of a highly efficient coding scheme, even if impractical for actual implementation, provides a new lower-bound for other bandwidth reduction schemes. Secondly, even if the implementation of a "hard-wired" configuration of the particular algorithm is not warranted, it may be valuable in the computer-to-computer communication environment.

Development of computer networks whose individual computer members may be separated by vast geographical distances is a modern concept which allows higher utilization of the modern "super" computers. The Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense network is an operational example, other networks are likely to follow. By design, a large scale computer network can perform arithmetic operations inexpensively. The data transmission, however, remains a relatively important cost factor. Image manipulation within the network will probably be expensive because of the requirement for large-volume data transmission. On the other hand, implementation of arithmetically complex coding/decoding algorithms may be easily programmed for the local "host" computers. The extra amount of computation may be offset to a significant degree by cost reduction for the transmission of the visual data.

Most of the picture coding techniques have considered mono-
chrome imagery only. Multispectral, color, and frame-to-frame
coding requirements have been addressed only very recently and
there is still a great deal of research needed in these areas.

Another objective of this dissertation is to extend the concepts
developed for monochrome imagery to the "third dimension"; specifi-
cally, color imagery and frame-to-frame redundancy are considered.

1.4   Overview of the Dissertation

Description of a research project on adaptive transform domain
coding is given in this dissertation. The presentation of the objec-
tives, development, and experimental results follow what the author
believes to be a carefully developed logical presentation which is
summarized in this section.

Chapter 1 is the Introduction and as such lays the groundwork
for the basic body of the dissertation. This chapter also places the
research project into perspective relative to the large amount of
research previously conducted in the field of picture coding. The
primary objectives of the dissertation are also spelled out in this
chapter.

Images are a specific class of signals and require careful con-
sideration if extreme redundancy reduction is desired. Chapter 2
addresses this important point of how images can be modeled and
characterized in terms of statistical and deterministic parameters.
Generation of the sampled image is considered, and important com-
parisons with the one-dimensional classical sampling theorem are

given. The image model as a statistical representation is also given. A review of the Fourier transform constraints is provided. Error criteria and image structure are briefly considered. The extremely important non-negative bound due to Lukosz is analyzed as related to sampling and relative importance of amplitude vs phase.

Chapter 3 presents the theoretical basis for the adaptive transform domain coding technique. It begins with the comparison of source and channel coding and consideration of schematic representation of adaptive techniques. Statistical properties of the Fourier and Walsh domain are analyzed. Phase and amplitude coding are considered in terms of quantization, sampling, and relative amount of information. Nonlinear effects of phase quantization are considered. Relative importance of phase is demonstrated via nonlinear filtering and gross reduction of amplitude information.

Chapter 4 is the first of three chapters discussing the experimental results. Monochrome image coding is considered in this chapter. Detailed discussion is given of the following topics: the algorithm, preprocessing, error analysis. Comparison is made with the conventional Markov model. Sensitivity analysis of noise effects on the coding algorithm is performed. Pictorial examples are included.

Experimental results of color coding are presented in Chapter 5. This chapter briefly reviews the theory of color perception and representation of color images. Extension of the monochrome algorithm of Chapter 3 is discussed and is followed by pictorial examples.

Frame-to-frame coding is considered in Chapter 6. Algorithm development is discussed and implementation includes both the Fourier and Walsh transforms. Pictorial results are provided. Unfortunately, the actual visual performance of the frame-to-frame coder can only be demonstrated in a realistic time-variant medium such as video presentation.

Chapter 7 summarizes the dissertation.

Appendix A contains the original test images. The numerical noise generation process of the large Fourier transform is considered in Appendix B.

2. IMAGE MODELING

The fundamental objective of the research project presented in this dissertation is the development of a very efficient source coding method for images. The emphasis is on the efficiency even at the expense of more complex algorithms and data handling. Clearly, the coder/decoder process must utilize as much a priori information as possible. The model should utilize both statistical and deterministic information.

This chapter addresses the role of image modeling in the image coding process.

2.1 Generation of the Discrete Image

Virtually all operations and transforms discussed in this disser-tation are performed numerically on discrete samples. It is tempting to follow the general approach to image coding and restrict the analysis to the discrete equivalent of the image. However, it should be remembered that images are generally viewed in analog form. The discretization of the image plays a fundamental part in the image coding process. In addition to the higher dimensionality of the prob-lem, there are very important factors that distinguish image sampling from sampling of one-dimensional time dependent signals. These concepts will now be considered.

Let the image be represented by $I = I(x, y)$, where x, y are spatial coordinates and I represents the analog image. The image is sampled on a square grid of lattice distance Δ. Let the sampled image be defined I_s.

15

The actual image sampling is always, almost by definition, performed by an optical system. The image (normally a photographic transparency, print, or an actual scene) sample of location x, y is imaged onto a photo detector whose output, ideally, is linearly proportional to the image brightness of that location.

The sample area can be considered via an aperture function $A(x, y)$. Typically $A(x, y)$ has the value of 1 in a small region around x, y and 0 elsewhere. Allowing for the finite aperture size, the sampled image has the following definition.

$$I_s(x, y) = comb\left(\frac{x}{\Delta}\right) comb\left(\frac{y}{\Delta}\right) I(x, y) * A(x, y)$$

$$= \iint_{-\infty}^{\infty} comb\left(\frac{p}{\Delta}\right) comb\left(\frac{s}{\Delta}\right) I(p, s) A(x-p, y-s) \, dpds \quad (2.1-1)$$

where

$$comb(x) = \sum_{n=-\infty}^{\infty} \delta(x-n)$$

and $\delta(x)$ is the Dirac delta function.

Considering the Fourier integral of this equation, one obtains (Goodman, 1968)

$$\widetilde{I}_s(u, v) = \left\{ comb(\Delta u) \, comb(\Delta v) * \widetilde{I}(u, v) \right\} \widetilde{A}(u, v) \quad (2.1-2)$$

The frequency domain coordinates are u, v and the symbol $\sim$ indicates the Fourier transform.

Equation (2. 1-2) is the classical result of sampling theory, however, one must be careful in its interpretation for image-related applications. The image I is always bandlimited. Visual scenes have structure at all levels, at the extreme, down to the micro or molecular structure. Permanent recordings do limit the spatial frequency extent and therefore become bandlimited. However, they introduce their own characteristic structure, for example, film grain. The bandlimiting is also performed by the optical system that performs the imaging.

The bracketed term in Equation 2. 1-2 indicates that the fundamental frequency band $\tilde{I}(u, v)$ is replicated at locations $n(1/\Delta)$, $m(1/\Delta)$, $n, m = 0, \pm 1, \pm 2, ..$ in the frequency plane. If twice the bandlimit of $\tilde{I}$ is larger than the sampling rate, $1/\Delta$, the replicated bands will partially overlap and undesirable aliasing occurs. The aperture function A should separate the fundamental band from its replicas. The requirement on A in this case is that

$$\tilde{A}(u, v) = 1; u, v \in \left[-\frac{1}{2\Delta}, \frac{1}{2\Delta} \right]$$

$$\tilde{A}(u, v) = 0 \quad \text{otherwise} \tag{2.1-3}$$

equivalently,

$$\tilde{A}(u, v) = \text{rect}(\Delta u) \, \text{rect}(\Delta v) \tag{2.1-4}$$

Equation (2.1-4) contradicts the physical constraint that the aperture function must be non-negative. Equation (2.1-4) leads to the unrealizable condition that

$$A(x, y) = \Delta^2 \, \text{sinc} \, \frac{x}{\Delta} \, \text{sinc} \, \frac{y}{\Delta} \qquad (2.1-5)$$

where

$$\text{sinc} \, x = \sin \pi x / \pi x.$$

A similar argument indicates (a more formal argument will be presented under the Lukosz bound section) that an optical system cannot perform the bandlimiting without attenuation in the band pass. On the other hand, the minimization of the attenuating effect of the optical system and/or the sampling aperture may lead to aliasing.

2.2 Statistical Consideration

The image sampling process and the non-negative image constraint are deterministic bounds. There are other descriptive constraints on images which can only be utilized through statistical consideration.

A wealth of knowledge has been developed in statistical communication theory and related disciplines which can be very useful in the design of image coding algorithms. The image can be considered as a sample function generated by a stochastic source. The statistics of the source may be available or can be estimated or, as is usually done, calculated from the image itself. For the latter case to be

valid, ergodicity for the calculated parameters usually must be assumed.

Knowledge of the image second order statistics can provide significant assistance in the development of efficient image coding algorithms. The image correlation function, $R(x_1, x_2; y_1, y_2)$ is defined as

$$R(x_1, x_2; y_1, y_2) = \overline{(I(x_1, y_1) - \overline{I(x_1, y_1)}) (I(x_2, y_2) - \overline{I(x_2, y_2)})}$$

(2.2-1)

The over-bar indicated ensemble averaging. I is the image which in this case is considered as a random process, and $I(x_i, y_j)$ is a sample of that process and is considered as a random variable.

The correlation function is usually estimated by involving the ergodicity argument and the assumption of wide sense stationarity. If R can be decomposed into the product of vertical and horizontal correlation functions; then

$$R(x, y) = R_x(x) R_y(y) \qquad (2.2-2)$$

The approximation of R_x and R_y by exponential function has been utilized for coding (Habibi and Wintz, 1971) as well as filtering (Pratt, 1972) of images, and for this case the correlation function is given by

$$R(x, y) = e^{-\alpha|x|} e^{-\beta|y|} \qquad (2.2-3)$$

Although the actual image coding techniques discussed in this dissertation do not utilize Equation (2.2-3), the discussion of this structure is desirable since many transform image coding algorithms are relying on the separable exponential form. In subsequent chapters comparison will be made between such approaches to image coding and the new algorithms presented in this dissertation. Some of the later analysis will require an explicit form for the correlation function; for example, the effect of additive noise on the coder, and the use of the exponential form because of its simplicity.

One should emphasize that Equation (2.2-1) refers to the recorded sampled image. The correlation properties of the analog visual scene are rarely available and can only be inferred from detailed knowledge of the sampling parameters.

The Fourier transform of Equation (2.2-1) is the conventional definition of power spectral density, $S(u, v)$. Using the aperture function A of subsection 2.1, it is straightforward to show that

$$S_s(u, v) = \left| \widetilde{A}(u, v) \right|^2 S(u, v) \qquad (2.2-4)$$

here, the subscript s denotes the sampled version. As indicated previously, the lack of precise knowledge of the sampling parameters does not permit accurate modeling of the original image. The structural form of Equation (2.2-4) permits a somewhat different interpretation. The sampled image can be considered as one which

has been processed by the linear spatial filter $\widetilde{A}(u, v)$. Consequently, the conventional PCM code available to the image coder reflects the essentially low-pass filtering effect of the sampling process.

2.3   Consideration of the Transform Domain

Image coding algorithms generally operate on image elements directly. The significant advances in digital hardware technology stimulated research in a new approach to image coding which have come to be known as transform coding. In this section, a short overview is given to the transform domain.

Let the image be denoted by I as in Chapter 1, $I = I(x, y, \lambda, t)$, indicating the functional dependence on the spatial coordinates (x, y), color (λ) and time (t).

A transform coder algorithm operates in a domain other than the original described by the four parameters: x, y, λ, t. The following symbolic representation can be written

$$\widetilde{I}(u_1, u_2, u_3, u_4) = T\left\{I(x, y, \lambda, t)\right\} \qquad (2.3-1)$$

T is the operator which performs the transformation between the two domains and it should be invertible. The latter requirement is due to the fact that without coding no ambiguity should be present in the image transformation. Consequently,

$$\widetilde{I}(x, y, \lambda, t) = T^{-1}\left\{I(u_1, u_2, u_3, u_4)\right\} \qquad (2.3-2)$$

and

$$TT^{-1} = T^{-1}T = \overline{I} \qquad (2.3-3)$$

where $\overline{I}$ is the identity operator.

Other than the requirement for invertability, T is completely general. Specifically, it may be linear or nonlinear. The operator T may be decomposable and it can operate on the continuous analog image or its sampled equivalent.

The choice of T is motivated by hope that $\tilde{I}(u_1, u_2, u_3, u_4)$ can be coded more efficiently.

Practical requirements restrict T to mathematical forms which are numerically implementable without excessive computation. The transform algorithms which have been successfully implemented can be grouped into three classes.

a)   Karhunen-Loéve (K-L) Transform

The image I is expanded into the eigenfunctions of the image covariance matrix. Although this transform is important from the theoretical viewpoint, its practical value is much less significant. The difficulties are lack of "fast" implementation, and in addition, the exact form of the covariance function usually is not available. In the presence of noise, the eigenfunction expansion will become degenerate. This is very significant and has not been considered in the context of image coding. The K-L transform emphasizes the second order image statistics. Its optimality is achieved for Gaussian processes which do not closely represent

images in general. The K-L transform assumes stationarity which is an additional assumption that is rarely met for typical images.

b)   Trigonometric Decomposition (Fourier Transform)

The image energy tends to concentrate for low frequencies, e.g., low values of u_1, u_2, u_3, u_4. These deterministic and statistical properties are useful to the transform coding algorithm and will be further considered in Chapters 3 and 4. The multidimensional Fourier transform is decomposable into a set of one-dimensional transforms and it can be implemented by the "fast" Fourier transform algorithm. The Fourier domain is also constrained by the Lukosz bound (subsection 2.5).

c)   Other Orthogonal Decompositions

Transform coding has also been successful in utilizing various fast orthogonal decompositions. The most well known among them is the Walsh transform. Although, no simple mathematical justification can be offered for their successful utilization, it can be shown that these functions are "approximately" trigonometric functions.

The particular value of the transforms under this category is their close similarity to the Fourier transform, however, they are suboptimal to it. What is meant by optimality in this case is deferred to the experimental chapters. In spite of this suboptimality, the non-trigonometric, orthogonal function decomposition may be preferred because of the ease of numerical implementation.

The Walsh decomposition can be accomplished without multiplication or division, and, consequently, its digital implementation is superior to that of the Fourier transform (Harmuth, 1972); although, this fact is more significant for smaller computers without hardware floating point multiply and divide registers.

Equations (2.3-1) and (.3-2) are implemented in numerical form; therefore, the discrete representation will be considered. If T is restricted to be a linear operator, these equations can be represented in (generalized) matrix notation.

$$\widetilde{I}(u_1, u_2, u_3, u_4) = \sum_x \sum_y \sum_\lambda \sum_t A(u_1, u_2, u_3, u_4, x, y, \lambda, t) \, I(x, y, \lambda, t)$$

(2.3-4)

In all practical cases, the multidimensional operator can be factored into a number of operators equal to the dimension of the problem. Let $A = A_1 A_2 A_3 A_4$ and equivalently $A = A_1(u, x) A_2(u_2, y) A_3(u_3, \lambda) A_4(u_4, t)$.

Specification of A_i, $i = 1, 2, 3, 4$ defines the transform and the numerical implementation. The following well-known representation exists for the discrete Fourier transform (Andrews, 1970)

$$A(u, x) = \left(\sqrt{N}\right)^{-1} \exp - \frac{2\pi i}{N} \, ux \qquad (2.3-5)$$

Similarly for the Walsh transform

$$A(u, x) = \left(\sqrt{N}\right)^{-1} (-1)^{\sum_{i=0}^{n-1} u_i x_i} \tag{2.3-6}$$

N is the order of matrix $A(u, x)$. It is arbitrary for the Fourier transform but restricted for the Walsh transform to values 2^n, where n is a positive integer. The variables x_i, u_i in Equation (2.3-6) are the binary representation of x and u respectively.

2.4 Error Criteria

Between the source and the destination, the image is subjected to significant processing. It is important to note again that the communication link of Figure 1.1-1 is digital and the source, the visual scene, is analog. It is highly desirable to quantify the image degradation due to the coding algorithm. Let I be the input to the coder and $\hat{I}$ its estimate at the destination. A measure of error, E may be schematically specified as a functional dependence G on the difference between I and $\hat{I}$,

$$E = G(I - \hat{I}) \tag{2.4-1}$$

with the constraint that $G(0) = 0$.

Although the practical implementation of Equation (2.4-1) is extremely useful, it is still an unsolved problem.

Determination of a useful error measure for image evaluation is extremely difficult because even the most approximate mathematical modeling of the human vision is available only in limited cases.

A conventional compromise to Equation (2.4-1) is the mean-square error between I and $\widehat{I}$, which can be written in terms of the previously-developed notation as

$$E = \iiiint\limits_{x\ y\ \lambda\ t} [I(x, y, \lambda, t) - \widehat{I}(x, y, \lambda, t)]^2 \, dxdyd\lambda dt$$

(2.4-2)

The equivalent form of Equation (2.4-2) for the discrete case is

$$E = \sum_x \sum_y \sum_\lambda \sum_t \left\{ I(x, y, \lambda, t) - \widehat{I}(x, y, \lambda, t) \right\}^2 \qquad (2.4-3)$$

The image energy, I_e, is obtained from the above two equations by letting $\widehat{I} = 0$. Consequently, the normalized mean square error as used in Chapters 4 through 6 is given by $100 \times E/I_e$ in percentages.

2.5   Non-Negative Bound (Lukosz)

The Fourier transform of a non-negative signal obeys various well-known constraints. Perhaps the most important is the amplitude constraint.

Let

$$G(u) = \int_{-\infty}^{\infty} g(x)\, e^{-2\pi j ux}\, dx \qquad (2.5\text{-}1)$$

if

$$g(x) \geq 0$$

then

$$\left| G(u) \right| \leq G(0) \qquad (2.5\text{-}2)$$

The inequality (2.5-2) is well known (Goodman, 1968). The very important extension of this inequality to bandlimited non-negative signals has unfortunately been relegated to obscurity. A properly sampled image does represent a non-negative band-limited signal and as such obeys the inequality discovered by W. Lukosz (Lukosz, 1962) and is designated in this dissertation as the Lukosz bound.

In his original paper Lukosz was concerned with the modulation transfer function properties of optical systems as related to incoherent imaging. The Fourier transform of the modulation transfer function (the point source image) of an optical system is non-negative and has an absolute cutoff frequency. Given this information, Lukosz intended to determine if any additional constraints are applicable beyond Equation (2.5-2). Structurally, the incoherent optical

transfer function and the Fourier transform of a band-limited non-negative image are equivalent; that is to say that by definition they satisfy the same requirements. Consequently, the mathematical derivation of the Lukosz bound is applicable to a band-limited image as well as to the optical transfer function.

The Lukosz bound can be derived for any number of dimensions. The bound becomes stronger with increasing numbers of dimensions. The mathematical derivation of this bound will be demonstrated in this section. For derivation of the two-dimensional case, the reader is referred to the original Lukosz paper.

Consider the Fourier transform paid as in Equation (2.6-1), with the additional constraint:

$$G(u) = 0, \text{ for } u \geq u_m \qquad (2.5-3)$$

where u_m is the cutoff frequency. Note also that Equation (2.5-2) is already applicable.

Let $h(x)$ be another non-negative function, not restricted to be band-limited. Clearly, the convolution of h and g is also non-negative.

$$h * g = \int_{-\infty}^{\infty} h(s) \, g(x - s) \, ds \geq 0 \qquad (2.5-4)$$

Assume h to be Fourier transformable,

$$H(u) = \int_{-\infty}^{\infty} h(x) \, e^{-2\pi jux} \, dx$$

Furthermore, it is easy to show that h $*$ g satisfies Equation (2.5-4) by utilizing the Fourier transform properties of the convolution integral. The previous statements become even more obvious in the framework of linear system theory, as Lukosz argued, where g represents a low-pass filter function and h is the input signal. However, the specific physical argument, while intuitively satisfying, is unnecessary to the mathematical derivation.

Let h(x) be the Dirac comb function, comb x/L, as defined previously in Equation (2.1-1). The comb x/L is a periodic function, where the period is L. Therefore, a Fourier series representation cf comb x/L exists, and it is (see also Figure 2.5-1 for the graphical demonstration):

$$\text{comb} \, \frac{x}{L} = 1 + 2 \sum_{n=1}^{\infty} \cos 2\pi nx/L \qquad (2.5-5)$$

Let

$$G(u) = |G(u)| \, \exp j\theta(u),$$

and let

$$1/L \geq u_m/2$$
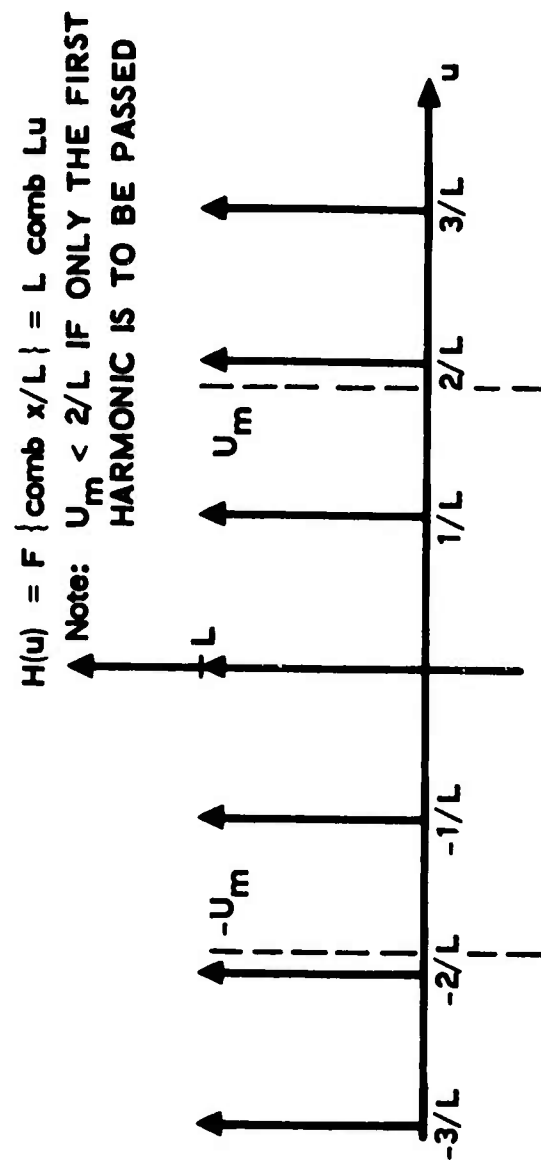
Figure 2.5-1. Demonstration of Equation (2.5-5)

then the convolution integral will preserve only the n = 1 term in Equation (2.5-5):

$$h * g = G(0) + 2\left|G(1/L)\right| \cos\left\{2\pi x/L + \theta(1/L)\right\} \quad (2.5\text{-}6)$$

Clearly, the inequality (2.5-2) is not sufficient to prevent the violation of inequality (2.5-4). The additional constraint must be imposed that

$$\left|G(1/L)\right| \le \tfrac{1}{2} G(0) \text{ for } 1/L \ge u_m/2 \quad (2.5\text{-}7)$$

Equation (2.5-7) is, in fact, the Lukosz bound for the region $u_m/2 \le u < u_m$. The derivation of other segments is based on choosing appropriate forms for h(x). Specifically, let h(x) have the following form

$$h(x) = \tfrac{1}{2}\left\{\text{comb }\frac{x - 1/8}{L} + \text{comb }\frac{x + 1/8}{L}\right\}$$

Equation (2.5-8) has the following Fourier series representation (see again Figure 2.5-2 for graphical demonstration).

$$h(x) = 1 + 2\sum_{n=1}^{\infty} \cos\tfrac{\pi}{4}n \cos 2\pi nx/L$$

$$= 1 + \sqrt{2}\cos(2\pi x)/L + 2\sum_{n=3}^{\infty} \cos\tfrac{\pi}{4}n \cos 2\pi L x/L$$

$$(2.5\text{-}8)$$

32

$H(u) = F\{comb \frac{1}{L}(x-1/8) + comb \frac{1}{L}(x+1/8)\}$

$= L \ comb \ Lu \ cos\pi Lu/4$

Note: $U_m < 3/L$ IF ONLY THE FIRST HARMONIC IS TO BE PASSED
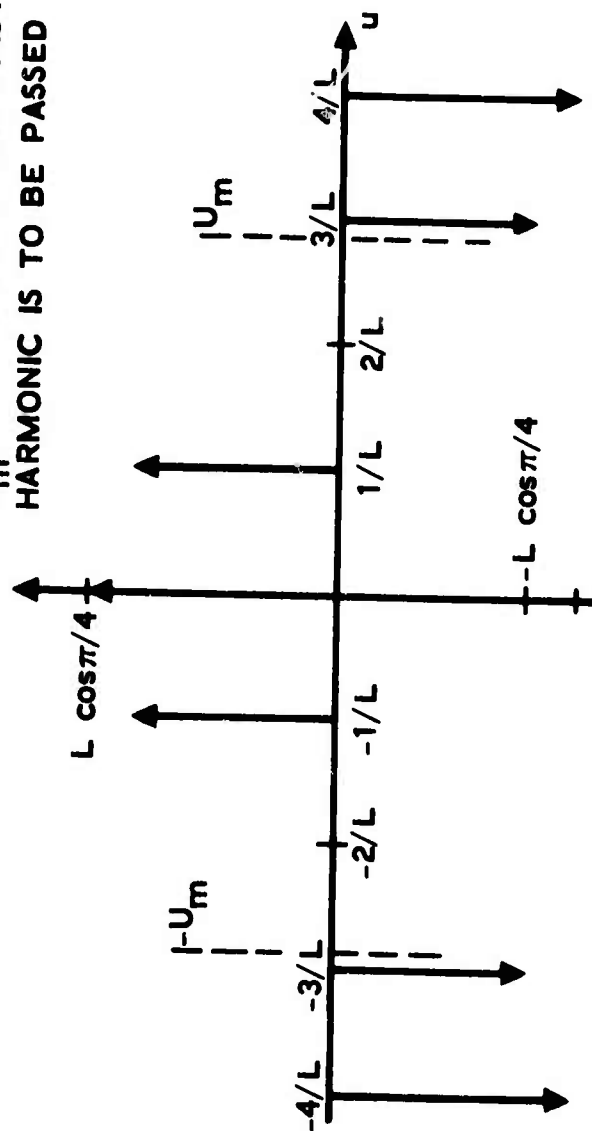
Figure 2.5-2. Demonstration of Equation (2.5-8)

For $1/L \geq u_m/3$, the form of $h * g$ is

$$h * g = G(0) + \sqrt{2} \left| G(1/L) \right| \cos \left\{ 2\pi x/L + \theta(1/L) \right\} \quad (2.5\text{-}9)$$

Since $h * g$ must not be negative,

$$\left| G(1/L) \right| \leq 1/\sqrt{2} \ G(0) \text{ for } 1/L \geq u_m/3 \quad (2.5\text{-}10)$$

Inequality expression (2.5-10) provides the next section of the Lukosz bound, namely, $u_m/3 \leq u < u_m/2$. It is equally valid for $u_m/2 \leq u < u_m$, but it is weaker than (2.5-7), therefore, not useful for that region.

The general form of the non-negative bound is obtained by choosing more complicated forms for $h(x)$. The general inequality is the following

$$\boxed{\left| G(u) \right| \leq G(0) \cos \frac{\pi}{n + 1} \text{ for } u_m/n \leq u \leq u_m/(n-1)}$$

$$(2.5\text{-}11)$$

and it is demonstrated in Figure 2.5-3. The argument u in inequality (2.5-11) is equivalent to $1/L$ in inequalities (2.5-7) and (2.5-10).

Inequality (2.5-11) is the Lukosz bound for one-dimensional, non-negative, band-limited signals. Its extension to higher dimensions can easily be obtained by successive Fourier decomposition

34
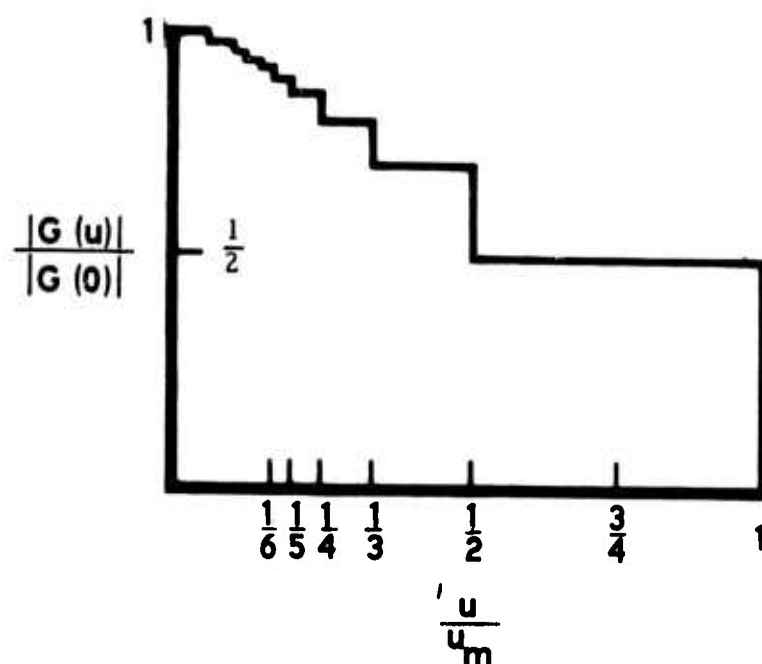
Figure 2.5-3.   One-Dimensional Lukosz Bound
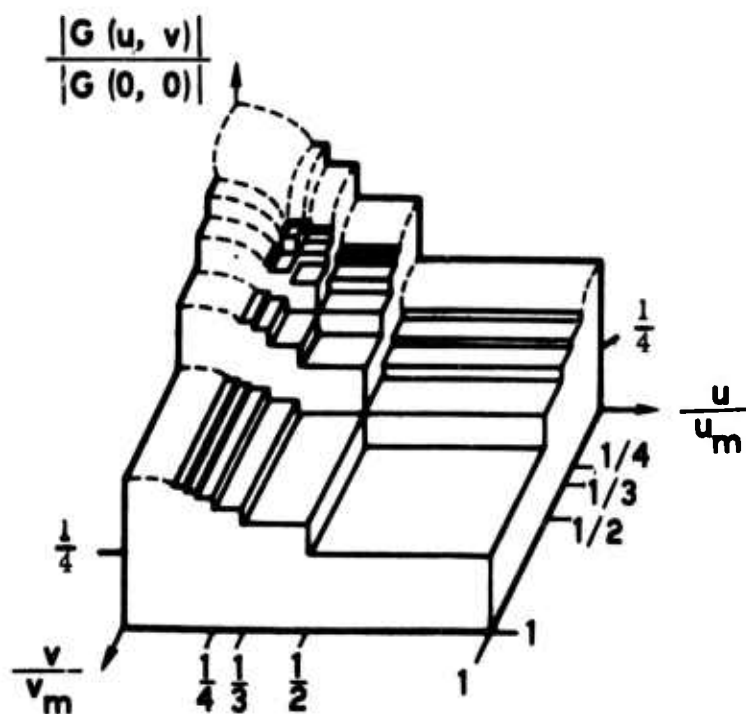
Figure 2.5-4.   Two-Dimensional Lukosz Bound

of the various dimensions. As previously stated, only the results will be given here.

Let g be non-negative, have two-dimensional Fourier transforms, G, and have band limits, u_m, v_m:

$$G(u, v) = \iint_{-\infty}^{\infty} g(x, y)\, e^{-2\pi j(xu + yv)}\, dx\, dy \qquad (2.5\text{-}12)$$

$$G(u, v) = 0, \text{ for } u \geq u_m \text{ or } v \geq v_m \qquad (2.5\text{-}13)$$

The functional form of the inequality for G is (Figure 2.5-4):

$$\frac{1}{2}\left\{\left|G(u, v)\right| + \left|G(-u, v)\right|\right\} = \frac{1}{2}\left\{\left|G(u, v)\right| + \left|G(u, -v)\right|\right\}$$

$$\leq G(0, 0) \cos \frac{\pi}{n+1} \cos \frac{\pi}{m+1}$$

$$(2.5\text{-}14)$$

for

$$u_m/n \leq u \leq u_m/(n-1)$$

and

$$v_m/n \leq v \leq v_m/(n-1)$$

The actual derivation (Lukosz, 1962) is straightforward although somewhat involved. By letting $G(u, v) = G_u(u)\, G_v(v)$ and

applying the one-dimensional bound to G_u and G_v individually the validity of Equation (2.5-14) was demonstrated by Lukosz. Note if G has directional symmetry, then

$$\frac{1}{2}\left\{\left|G(u,v)\right| + \left|G(-u,v)\right|\right\} = \frac{1}{2}\left\{\left|G(u,v)\right| + \left|G(u,-v)\right|\right\} = \left|G(u,v)\right|$$

(2.5-15)

Equation (2.5-15) can easily be proven by the well-known property of the Fourier transform of real functions in which G obeys:

$$G(u,v) = G^*(-u,-v) \qquad (2.5-16)$$

It easily follows that

$$\left|G(u,v)\right| = \left|G(-u,-v)\right| \qquad (2.5-17)$$

and

$$\left|G(-u,v)\right| = \left|G(u,-v)\right| \qquad (2.5-18)$$

It can easily be shown via Equation (2.5-18) that if $\left|G\right|$ is symmetric around the u axis, it has symmetry around the other axis as well.

Before proceeding to the derivation of additional constraints based on inequality (2.5-14), a few general comments on the importance of this inequality are in order.

The Lukosz bound restricts the amplitude range in the Fourier domain, it does not, however, constrain the values the phase may assume. One can qualitatively argue that in some sense the phase

carries more "information" about the non-negative sampled image than the amplitude. This statement, which will later be considered in a more formal presentation, is quite significant for the various areas of image processing, including holography, where, in fact, the superiority of phase information has been observed experimentally (Kermisch, 1970).

Actually the inequalities (2.5-11) and 2.5-14) can further be strengthened. The average values of Figures 2.5-3 and 2.5-4 are clearly larger than 1/2 and 1/4, respectively. It can, however, be shown, and again the reader is referred to the original paper for the derivation, that, 1/2 G(0) and 1/4 G(0,0) are the appropriate limits for the one- and two-dimensional cases, respectively. For the one-dimensional case

$$\int_0^{u_m} \left| G(u) \right| du \leq \frac{1}{2} G(0) \qquad (2.5\text{-}19)$$

and for the two-dimensional case

$$\int_0^{u_m} \int_0^{v_m} \left| G(u,v) \right| dudv \leq \frac{1}{4} G(0,0) \qquad (2.5\text{-}20)$$

The implication of Equations (2.5-19) and (2.5-20) is that for no image, can G actually assume the upper bound in the Fourier

domain. The functions which, in fact, satisfy Equations (2.5-19) and (2.5-20) with equality are for the one-dimensional case:

$$\left|G(u)\right| = G(0)\left[1 - \frac{|u|}{u_m}\right], \; |u| \le u_m \qquad (2.5-21)$$

For the two-dimensional case:

$$\left|G(u,v)\right| = G(0,0)\left[1 - \frac{|u|}{u_m}\right]\left[1 - \frac{|v|}{v_m}\right], \; |u| \le u_m, \; |v| \le v_m$$
$$(2.5-22)$$

It is interesting to note that G reaches the bound at a single point:

$$G\left(\frac{1}{2}u_m\right) = \frac{1}{2}G(0), \text{ and}$$

$$G\left(\frac{1}{2}u_m, \frac{1}{2}v_m\right) = \frac{1}{4}G(0,0).$$

Except for this point, G as defined in Equations (2.5-21) and (2.5-22) lies below the appropriate non-negative limit. The two special functions, (2.5-21) and (2.5-22), represent for the optical case the modulation transfer function for the uniformly lit slit and rectangular aperture, respectively.

The various inequalities (2.5-7), (2.5-14), (2.5-19), and (2.5-20), allow an information theoretic interpretation of the Fourier domain for non-negative signals.

The entropy associated with an image is invariant under the Fourier transform as well as any other transform for which the

Jacobian of transformation is unity. If no a priori information is available, the image entropy is uniformly distributed in the frequency domain by assumption. This type of reasoning yields upper bounds on the entropy rather than entropy estimates for actual images whose correlation properties are known. Given, thus, that the image entropy is divided between amplitude and phase, it is important to learn what affects the constraints (2.5-7), (2.5-14), (2.5-19), and (2.5-20) will have on the entropy division. Assumption of no a priori information implies, on the basis of Equation (2.5-2) alone, that the Fourier domain represents a uniform entropy density for spatial frequencies below the band limit. Restriction of the allowed amplitude range will proportionally limit the entropy. The ratio of the entropies with and without the Lukosz bound is 1/2 and 1/4 for the one-dimensional and two-dimensional cases, respectively. This statement follows from the inequalities (2.5-19) and (2.5-20). One can argue that, for band-limited, non-negative images, the entropy associated with the phase is larger by a factor of 2 and 4 for the one- and two-dimensional cases, respectively.

The optical analog is the case of incoherent imaging, for which it can be argued, as Lukosz did, that the optical system by virtue of its low-pass filtering will limit the information transfer by 1/2 and 1/4 for one and two dimensions, respectively.

The Lukosz bound is a significant contribution to the science of the signal processing of non-negative band-limited signals. The

implication of the importance of phase over amplitude in digital image processing is useful information and has strongly motivated the research in this dissertation.

3. IMAGE SOURCE CODING

The transmission of data consists of two distinct coding steps: source coding and channel coding. Schematic representation of the classical communication problem was reviewed in Chapter 1. This dissertation treats the image coding problem as one which fits into the domain of source coding. This approach permits structural separation of image coding from the consideration of channel errors.

In this chapter various aspects of image coding are considered. The basic theme of the dissertation is that the phase (yet to be explicitly defined) is the primary parameter whose fidelity should be maintained in the coding process. The various steps that constitute the coding process are considered in the context of phase coding. The primary transform domain is that of the Fourier, however, extension is made to the Walsh domain as well. In fact, successful utilization of phase in other than the Fourier domain is a discovery which, prior to this dissertation, has not appeared in the literature as far as the author is aware.

3.1 Statistics of the Fourier Transform

The various coding schemes of Chapters 4 through 6 utilize the properties of the transform domain. The primary transform is the Fourier which has extremely advantageous properties from the coding standpoint. The close similarity between the Fourier and Walsh decompositions makes the latter transform also useful. The statistical properties of the Fourier transform domain are explored in this section, the extension to the Walsh domain is the topic of the next section.

Let $\tilde{I}(\underline{x})$ and $\tilde{I}(\underline{u})$ be a Fourier transform pair. To simplify notation, the image coordinates are condensed into vector form. Vectors $\underline{u}$, $\underline{x}$ have a number of components equal to the dimension of the coding problem. The monochrome problem has two dimensions for this case,

$$\underline{u} = \{u, v\}$$

$$\underline{x} = \{x, y\}$$

The frame-to-frame, or color coding problem is of three dimensions, for this case.

$$\underline{u} = \{u, v, w\}$$

$$\underline{x} = \{x, y, t\}$$

The vector notation permits statistical analysis of the Fourier transform of an image without specification of the dimension.

In addition, the infinite extent of the image plane implies that the Fourier domain is uncorrelated in the limit as the number of samples grows to infinity (Davenport and Root, 1958, Section 6-4).

The functional form of the power spectral density is required, if quantization of the transform samples is to be accomplished efficiently. All transform coding techniques require an estimate of the power spectral density, their overall performance is largely determined by how well the power spectral density estimation is accomplished.

Information-theoretic discussion of the frequency plane, based on the Lukosz bound, already implied a certain superiority of the phase. Stochastic consideration of the Fourier domain allows additional interpretation, in fact, a general definition of the phase. This dissertation expands the phase concept to what will be referred to as the unconventional definition.

a) Conventional Definition

The complex valued function $\widetilde{I}$ is the sum of real and imaginary components,

$$\widetilde{I}(\underline{u}) = \widetilde{I}_R(\underline{u}) + j\,\widetilde{I}_I(\underline{u})$$

the phase $\theta(\underline{u})$ associated with $\underline{u}$ is normally defined as

$$\theta(\underline{u}) = \tan^{-1}\widetilde{I}_I(\underline{u})/\widetilde{I}_R(\underline{u}) \qquad (3.1-1)$$

The definition in Equation (3.1-1) is required if the various well-known phase-related deterministic properties of the Fourier transform are to be utilized.

b) Unconventional Definition

Under the assumption, based on experimental evidence, that $\widetilde{I}_R$ and $\widetilde{I}_I$ are approximately Gaussian, θ is uniformly distributed and uncorrelated for different values of $\underline{u}$, that is

$$E\{\theta(\underline{u}_1),\ \theta(\underline{u}_2)\} = 2/3\ \pi^3\ \delta(\underline{u}_1 - \underline{u}_2) \qquad (3.1-2)$$

and

$$E\left\{\left|I_I(\underline{u})\right|^2\right\} = E\left\{\left|I_R(\underline{u})\right|^2\right\} = \frac{1}{2}\,S(\underline{u}) \qquad (3.-13)$$

In most practical situations $S(\underline{u})$ is a smooth surface, which means that $S(\underline{u}_1) \approx S(\underline{u}_2)$ for $\left|\underline{u}_1 - \underline{u}_2\right| < M$. The expression $\left|\underline{u}_1 - \underline{u}_2\right|$ is the Euclidean distance for vectors $\underline{u}_1$ and $\underline{u}_2$. For the sampled case, a reasonable value for M might be at least 5 (in harmonics). The comment should be interjected that the I can be only approximately Gaussian since its components are restricted in range by the D.C. term and for the band-limited case by the additional Lukosz bound.

Based on the smoothness of S, the following stochastic unconventional phase definitions can be made, with the previously-made restriction $\left|\underline{u}_1 - \underline{u}_2\right| < M$.

$$\theta(\underline{u}_1,\ \underline{u}_2) = \tan^{-1}\left\{I_K(\underline{u}_1)/I_L(\underline{u}_2)\right\} \qquad (3.1-4)$$

Subscripts K and L represent the actual independent subscript assignments from I and R (imaginary and real) if $\underline{u}_1 \neq \underline{u}_2$. K and L represent different subscripts if $\underline{u}_1 = \underline{u}_2$. The following forms for θ are allowed under the unconventional definition

$$\theta(\underline{u}_1, \ \underline{u}_2) = \tan^{-1}\left\{ I_R(\underline{u}_1)/I_R(\underline{u}_2) \right\}, \ \underline{u}_1 \neq \underline{u}_2$$

$$\theta(\underline{u}_1, \ \underline{u}_2) = \tan^{-1}\left\{ I_R(\underline{u}_1)/I_I(\underline{u}_1) \right\}$$

The following definition is not permitted

$$\theta(\underline{u}_1, \ \underline{u}_1) = \tan^{-1}\left\{ I_R(\underline{u}_1)/I_R(\underline{u}_1) \right\} = \frac{\pi}{4}$$

since, in this case, θ is a single value rather than a random variable. The stochastic phase definition is important because it gives validity to phase coding in domains other than the Fourier. Experimental demonstration of the utility of the stochastic phase will be given in this chapter.

3.2   Extension to the Walsh Domain

The Fourier transform of an image tends to be uncorrelated. The existence of uncorrelated samples permitted definition of the generalized phase. Although the Fourier transform is unique in having the above-mentioned properties, other linear transformations may approximate the Fourier transform in some sense. One specific implementation will involve the Walsh functions.

Let f_k be an element of an N component vector (that is $K = 1, 2, \ldots N$). Two distinct transforms of f_k and a_i and b_j which are also elements of N component vectors, therefore

$$a_i = \sum_k G_{ik} f_k \qquad (3.2-1)$$

$$b_i = \sum_k H_{ik} f_k \qquad (3.2-2)$$

where matrices G and H are invertible matrices of order N. The summation in Equations (3.2-1) and (3.2-2) is over N components, the same convention will remain in force for the rest of this section.

Although, Equations (3.2-1) and (3.2-2) can represent any linear decomposition, the specific assignment will be made where G will represent the Walsh and H the Fourier decomposition.

By straightforward manipulation, it can be shown that

$$f_k = \sum_j G_{kj}^{-1} a_j = \sum_j H_{kj}^{-1} b \qquad (3.2-3)$$

and, therefore,

$$a_1 = \sum_j \sum_k G_{1k} H_{kj}^{-1} b_j \qquad (3.2-4)$$

The following definition is introduced for notational convenience

$$Z_{1j} = \sum_k G_{jk} H_{kj}^{-1} \qquad (3.2-5)$$

Consequently, the transform values are related through the linear relationship:

$$a_\ell = \sum_j Z_{\ell j} b_j \qquad\qquad (3.2\text{-}6)$$

Let $\left\{g_i(x_k)\right\}$ and $\left\{h_i(x_k)\right\}$ be orthonormal basis vectors generating the space in which f_k is defined. Note that $g_i(x_k)$ is the k-th element of the i-th vector. Obviously, both i and k have index values 1 through N.

For the special case where $G_{ki} = g_i(x_k)$ and $H_{ki} = h_i(x_k)$, it is easy to demonstrate that $G_{ik} = G_{ki}^{-1}$ and $H_{ik}^{*} = H_{ki}^{-1}$.

It is desirable to treat the a_i's and b_i's as zero mean real random variables and consider transformation of the second order statistics. Clearly,

$$E\left\{a_\ell \, a_n\right\} = \sum_j \sum_k Z_{\ell j} Z_{nk} E\left\{b_j b_k\right\} \qquad\qquad (3.2\text{-}7)$$

If the Fourier designation is given to H_{ik}, then, according to the results of the previous section,

$$E\left\{b_j \, b_k\right\} = \sigma_j^2 \, \delta_{jk} \qquad\qquad (3.2\text{-}8)$$

where

$$\delta_{jk} = 1 \qquad j = k$$

$$= 0 \qquad j \neq k$$

Consequently,

$$E\left\{a_\ell\, a_n\right\} = \sum_j Z_{\ell j} Z_{nj} \delta_j^2 \qquad (3.2-9)$$

$$E\left\{a_\ell^2\right\} = \sum_j Z_{\ell j}^2 \delta_j^2 \qquad (3.2-10)$$

The previous section indicated that Fourier transform samples have Gaussian distribution. By Equation (3.2-6), it is observed that the a_j's also tend toward a Gaussian stochastic process. If the choice for G_{ik} is such $E\left\{a_\ell a_n\right\} \simeq \delta_{\ell k} E\left\{a_\ell^2\right\}$, that is, the a_ℓ's are also uncorrelated, one can define amplitude and phase on pairs of random variables, say a_ℓ, a_n.

If $E\left\{a_\ell^2\right\} = E\left\{a_n^2\right\}$, the functional form of appropriate probability density functions for the amplitude and phase should be the same as the ones defined for the Fourier transform.

The specification of G for the Walsh decomposition can be written in terms of the appropriate orthonormal basis vectors. Utilizing the conventional notation (Harmuth, 1972)

$$G_{ki} = \text{wal } (i,\ x_k) \qquad (3.2-11)$$

Walsh functions can be generated through the following difference equation.

$$\text{wal } (2j + P, \ x) =$$

$$(-1)^{[j/2]+P} \left\{ \text{wal} \left[j, \ 2\left(x + \frac{1}{4}\right) \right] + (-1)^{j+P} \text{ wal} \left[j, \left(2(x - \frac{1}{4})\right) \right] \right\}$$

$$(3.2-12)$$

$$P = 0 \text{ or } 1; \qquad j = 0, 1, 2, \cdots$$

$$\text{wal } (0, \ x) = 1 \text{ for } -\frac{1}{2} \leq x \leq \frac{1}{2}$$

$$= 0 \text{ otherwise}$$

and $[j/2]$ is the largest integer smaller than or equal to $j/2$.

The Z matrix can be generated by decomposing each Walsh function into a Fourier series. Walsh functions have similar symmetry as the sines and cosines. Denoting the even and odd Walsh functions as cal and sal, respectively, it follows

$$\text{wal } (2i, \ x) = \text{cal } (i, \ x)$$

$$\text{wal } (2i - 1, \ x) = \text{sal } (i, \ x) \qquad (3.2-13)$$

As previously indicated, real Fourier decomposition is where the basis functions, h_i's are sine and cosine functions, similarly the g_i's are the cal and sal functions. Because of the even-odd symmetry of both sets of basis vectors, even functions of one set can be represented by only even functions of the other set. Similar representation holds for odd basis vectors. The same symmetry results in the following restriction for the Z matrix.

$$Z_{\ell k} = 0 \text{ for } |\ell - k| = \text{odd integer} \qquad (3.2-14)$$

The specific result of Equation (3.2-14) is that elements in the Walsh domain which are separated by odd-number elements will be uncorrelated. The choice of adjacent element pairs for amplitude and phase specification is strongly motivated by the symmetry consideration.

Although, simple functional form does not exist for the Z matrix, numerical generation of the elements can easily be performed for specific transform pairs. As an example, consider the Walsh into Fourier decomposition for $N = 1024$ values. For a specific choice of ℓ, e.g., the ℓ-th Walsh function, ℓ-th row of the Z matrix is generated. The inverse of Z is similarly generated by the decomposition of particular sine and cosine functions into Walsh functions. Numerical examples are shown in Figures 3.2-1 through 3.2-8. These figures indicate the recognized similarity between the Walsh and trigonometric functions. It is interesting to observe that diagonal elements of Z dominate each row.

For completeness, the "fast" computability of the Walsh and Fourier transforms should be pointed out. The straightforward application of Equations (3.2-1) and 3.2-2) requires N^2 operation (operation $\triangleq$ one complex multiplication for Fourier and $\triangleq$ one addition or one subtraction for Walsh). The particular form of G and H permits a much more rapid implementation of these transforms where the number of operations is reduced to N log N (Andrews, 1970; Harmuth, 1972; Cooley and Tukey, 1965).
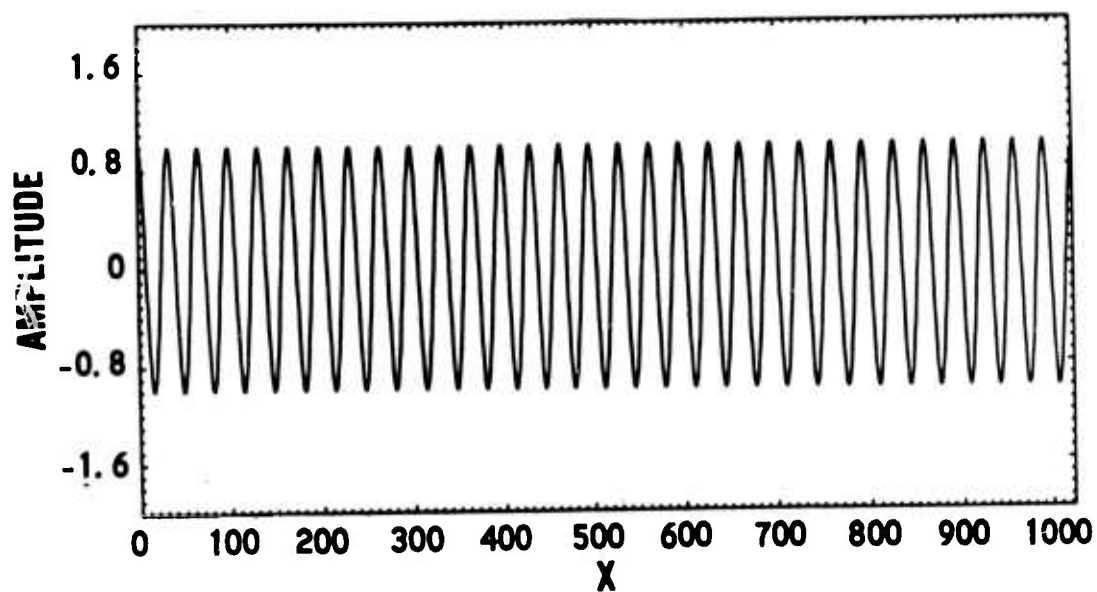
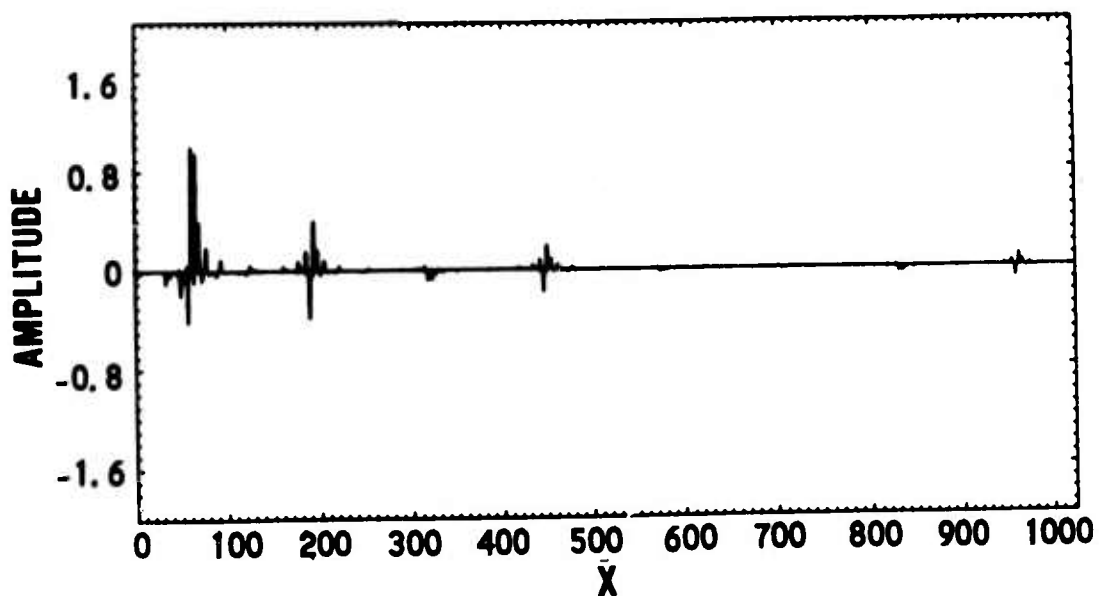Figure 3.2-1.   Sixty-Third Trigonometric Function

Figure 3.2-2.   Walsh Decomposition of Sixty-Third
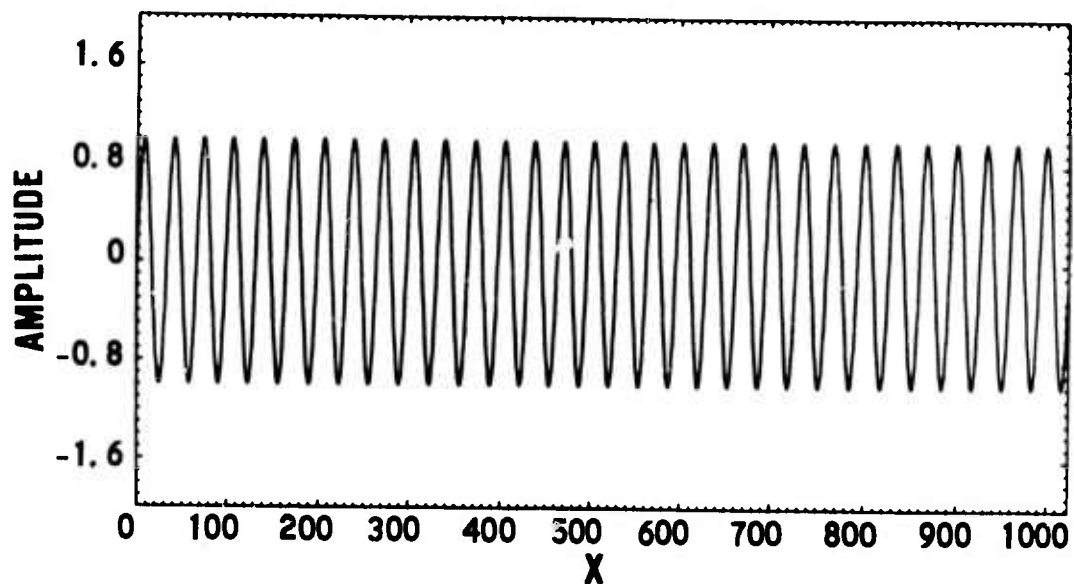Trigonometric Function

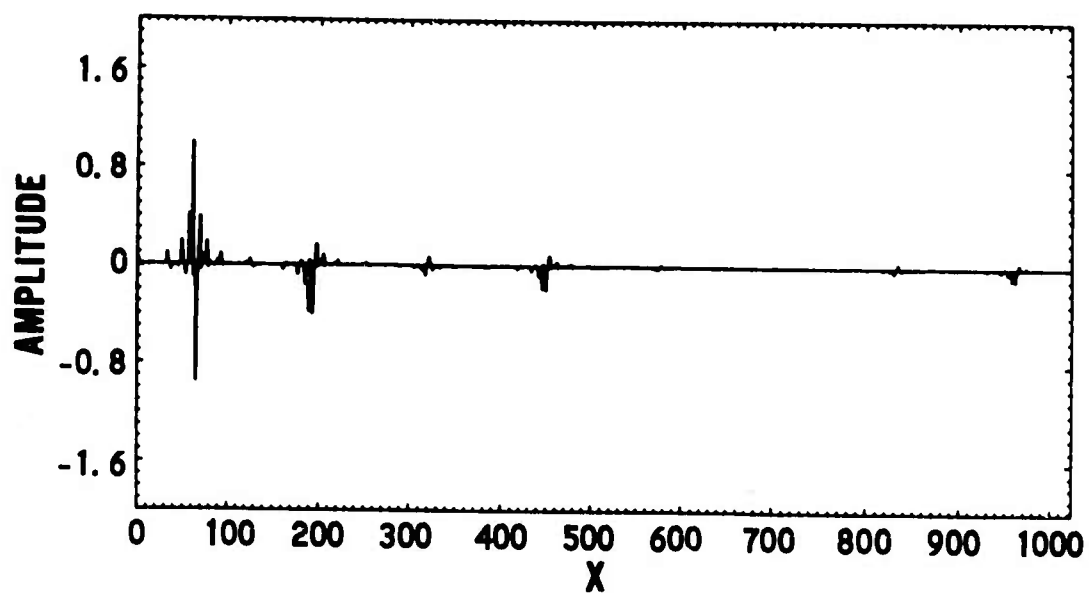Figure 3.2-3.   Sixty-Fourth Trigonometric Function

Figure 3.2-4.   Walsh Decomposition of Sixty-Fourth
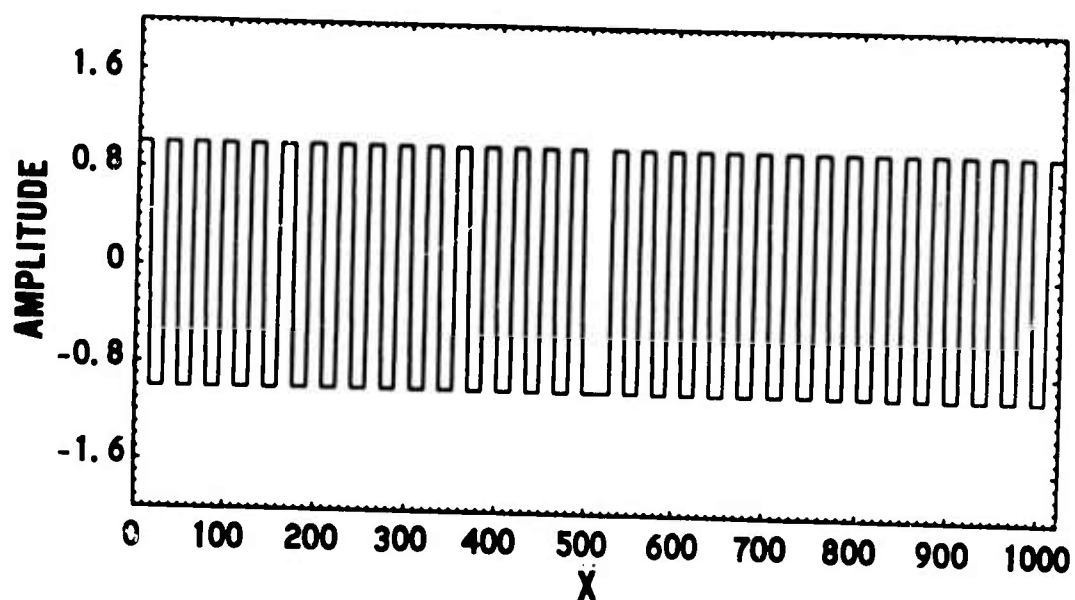Trigonometric Function

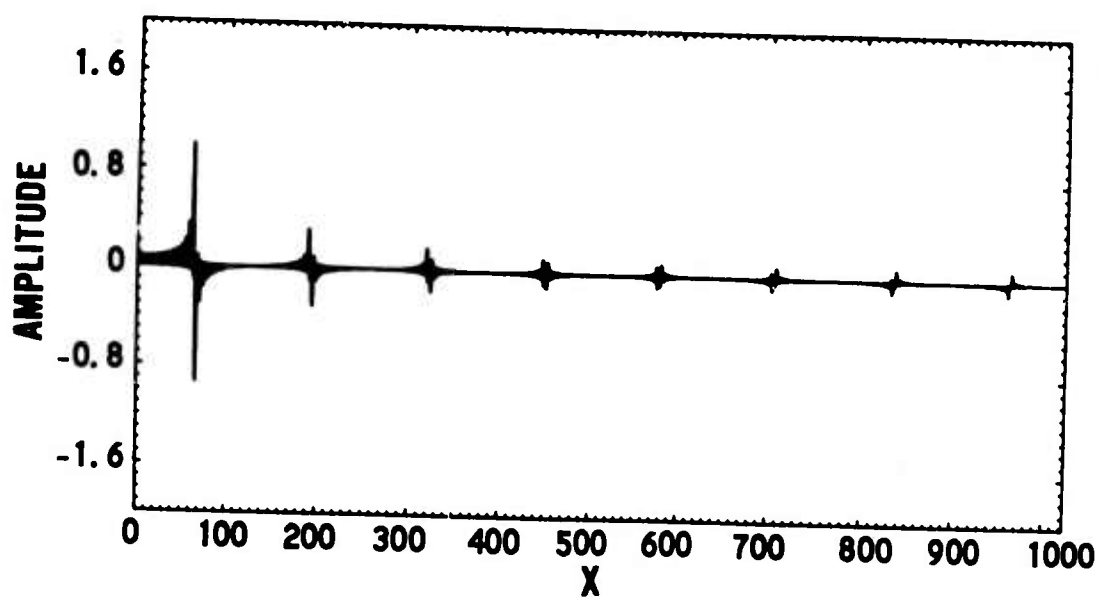Figure 3.2-5. Sixty-Third Walsh Function

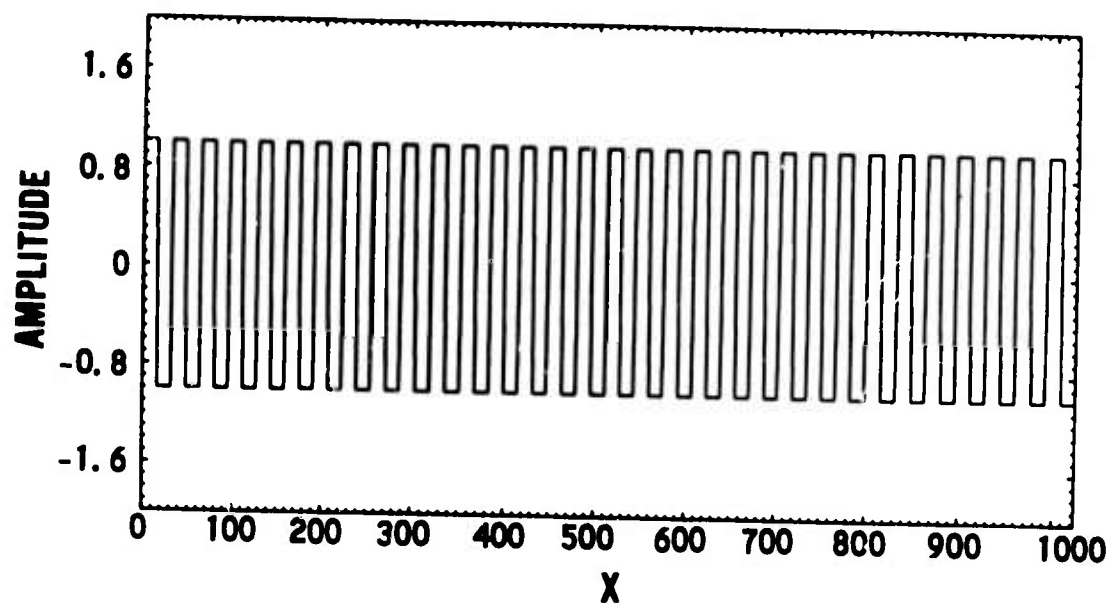Figure 3.2-6. Fourier Decomposition of Sixty-Third
Walsh Function

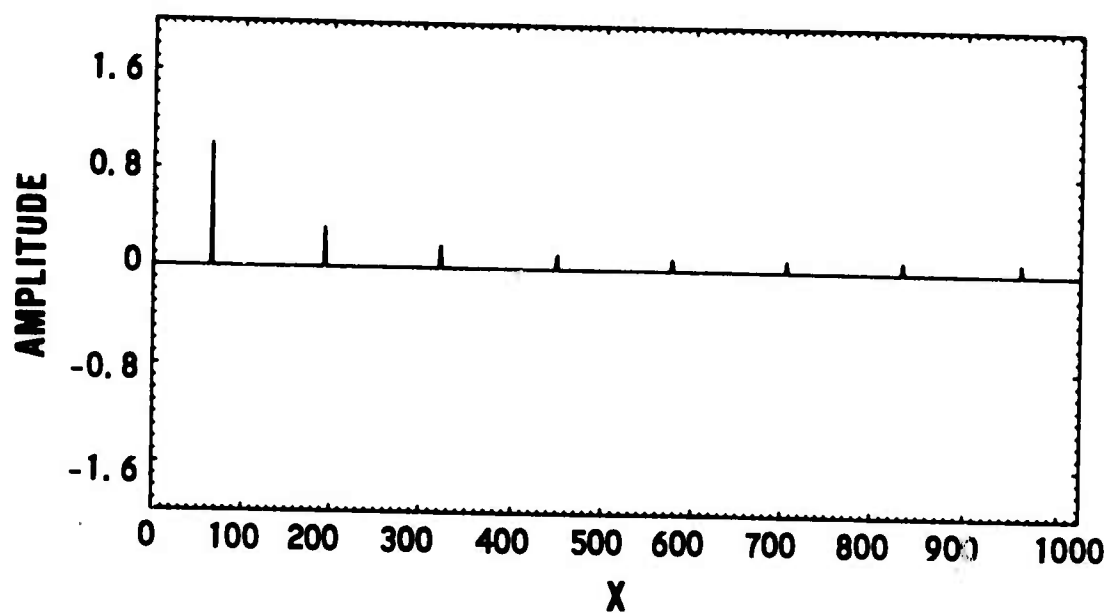Figure 3.2-7.   Sixty-Fourth Walsh Function

Figure 3.2-8.   Fourier Decomposition of Sixty-Fourth
Walsh Function

The "fast" algorithms are important for efficient coding implementation. Particularly for large data blocks, the efficiency factor N/log N can be significant. The "fast" algorithm is available for the Z matrix as well and it was utilized for the generation of Figures 3.2-1 through 3.2-8.

3.3 Quantization

The continuous image parameters must be expressed in discrete, that is to say quantized, form before numerical operation on them can be performed. Formally, quantization is equivalent to a noninvertible mapping of the real numbers onto a finite set of integers. It is also equivalent to a one-to-one mapping of finite or semiinfinite sections of the real axis to a finite set of integer numbers.

According to the last definition, each member of a set is assigned an integer designation. All members of a set are assigned the same integer assignment. Conversely, given a particular integer assignment, no unique determination of the original real value can be made.

It is obviously imperative to optimize the appropriate quantization procedures. This step involves the selection of the optimum quantization rules, based on the statistical model of the parameter to be quantized.

The discretization of a continuous parameter always results in a permanent, hopefully negligible, distortion. This distortion may appear as an effective noise term or an actual structural distortion.

For the first case, the number of quantization levels are large and the appropriate effects can be modeled by additive white noise. The second case occurs for coarse quantization, for which the nonlinear aspect of quantization dominates.

The following basic model will be considered. Let x be a continuous random variable with a probability density function $P(x)$. The functional form of the quantization can be expressed in terms of the previously-introduced rect function as

$$Q(x) = \sum_{j=1}^{N} \hat{x}_j \; \text{rect}\left[\frac{1}{x_j - x_{j-1}} \left(x - \frac{1}{2}(x_j + x_{j-1})\right)\right] \quad (3.3\text{-}1)$$

In Equation (3.3-1) there are N integer assignments. To each integer another real value, $\hat{x}_j$ is assigned. The $\hat{x}_j$ is the reconstruction value or the estimate of x. The specification of the parameters in Equation (3.3-1) should be such that $\hat{x}_j$ should closely "approximate" x. If the mean-squared error (MSE) is the performance measure, then

$$\text{Error} = \min\left\{\int P(x) \, (Q(x) - x)^2 \, dx\right\} \quad (3.3\text{-}2)$$

Where minimization is performed over all x_j's and $\hat{x}_j$'s for a given N. The solution of Equation (3.3-2) is well known (Max, 1960); it is

$$x_j = \frac{1}{2}(\hat{x}_j - \hat{x}_{j-1}), \; j = 2, \ldots N \quad (3.3\text{-}3)$$

and

$$\int_{x_j}^{x_{j+1}} (x - \widehat{x}_j)\, P(x)\, dx = 0, \quad j = 1, 2, \ldots N \qquad (3.3-4)$$

Equations 3.3-3) and (3.3-4) can be solved by iterative techniques for given density functions. A note of caution should be interjected. Equations (3.3-3) and (3.3-4) are formal solutions <u>given</u> the $P(x)$. In image coding, the relevant parameters are themselves estimated. Utilization of an erroneous model may result in a poor quantization procedure even though the solutions in Equations (3.3-3) and (3.3-4) are faithfully followed.

If $P(x)$ is uniform over a finite region, say $[x_0, x_N]$, Equation (3.3-1) becomes the uniform quantizer.

$$Q_u(x) = \sum_{j=1}^{N} \frac{(j-1)}{N} (x_N - x_0)\, \text{rect}\left\{\frac{N}{X_N - x_0}\left(x - \left(\frac{j-1}{N}\right)(x_N - x_0)\right)\right\}$$

$$(3.3-5)$$

Another often-used quantization strategy, known as companding (Smith, 1957) involves a two-stage process. First x is mapped into y, $y = f(x)$, which is random variable uniformly distributed between $[0, 1]$. The random variable y is operated on by the uniform quantizer. The reconstruction levels of x and y are determined by the inverse mapping, f^{-1}, $(\widehat{x}_j = f^{-1}\widehat{y}_j)$. The mapping is the distribution function of x:

$$y = f(x) = \int_{-\infty}^{x} P(u) \, du \qquad (3.3-6)$$

The reconstruction levels x_j and y_j are uniquely related by the one-to-one mapping, f. By construction, y_j occurs with equal probability, thus, this case corresponds to the maximum entropy the quantized values may have. This latter type of quantizer procedure is suboptimal when MSE is the performance criterion; however, for numerous density functions, optimum performance is closely approached.

Quantization schemes can be closely approximated by simplified procedures for fine quantization (Panter and Dite, 1951). The coding schemes of Chapters 4 through 6 involve coarse quantization in the transform domain, thus, these procedures are not relevant and will not be further explored.

3.4   Amplitude vs Phase Quantization Effects

The underlying theme of this dissertation is the superiority of phase information. It is particularly relevant to consider distortions introduced by the quantizing process. In this subsection, the generalized phase and amplitude will be considered. The assumption is made that application of the image transform (Fourier or Walsh) results in uncorrelated samples. Amplitude and phase are defined over pairs of values as in subsection 3.3 under the unconventional definition.

Let θ and r be a phase and amplitude pair where θ is uniformly distributed in $[-\pi, \pi]$ and r has Rayleigh distribution (Thomas, 1968, Chapter 4). The following procedure will be implemented. Amplitude and phase will be independently quantized, one at a time, and the appropriate MSE generated will be compared.

a)   Phase Quantization

The uniform quantizer is optimum for the phase. The actual error in the N level quantization process of a single phase value in one of the N regions, say $[0, 2\pi/N]$, is

$$\text{Error} = A^2 \left| e^{i\theta} - e^{-i\frac{\pi}{N}} \right|^2 \qquad (3.4\text{-}1)$$

The A^2 is the energy associated with random variable r. Mean-squared phase error (MSE) is obtained by averaging Equation (3.4-1) over θ and all N quantization regions. Because of the symmetry in θ, each of the quantizing sections is statistically equivalent, therefore

$$MSE = A^2 \ E \left| e^{j\theta} - e^{j\frac{\pi}{N}} \right|^2$$

$$= 2A^2 \left\{ 1 - E \cos\left(\theta - \frac{\pi}{N}\right) \right\}$$

$$= 2A^2 \left\{ 1 - \frac{N}{2\pi} \int_0^{\pi/N} \cos\left(\theta - \frac{\pi}{N}\right) d\theta \right\} \quad (3.4-2)$$

$$= 2A^2 \left\{ 1 - \frac{N}{\pi} \sin\frac{\pi}{N} \right\}$$

The more meaningful expression is the relative mean-squared error (RMSE) or, equivalently, where r is assumed to have unit energy,

$$RMSE = 2 \left\{ 1 - \frac{N}{\pi} \sin\frac{\pi}{N} \right\} \quad (3.4-3)$$

The approximate form of Equation (3.4-3) for large values of N is

$$RMSE \ (N \ is \ large) \approx \frac{1}{3}\left(\frac{\pi}{N}\right)^2$$

Quadratic dependence of the MSE on N is typical for all quantization procedures for large N.

b)   Amplitude Quantization

Amplitude is assumed to be Rayleigh-distributed,

$$P(r) = \frac{r}{\sigma^2} \ e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}, \ r < 0 \quad (3.4-4)$$

Maximum entropy quantization (companding) of sub-section 3.6 will be utilized (Habibi, 1973). The function f is required, which is the appropriate distribution function:

$$f(r) = \int_0^r \frac{s}{\sigma^2} e^{-\frac{1}{2}\left(\frac{s}{\sigma}\right)^2} ds$$

$$= 1 - e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2} \tag{3.4-5}$$

The inverse of f is also available in closed form of

$$f^{-1}(s) = \sigma \sqrt{-2 \log (1 - u)} \tag{3.4-6}$$

Let $\sigma = 1$; the RMSE for the Rayleigh process using the formalism of subsection 3.3 is

$$RMSE = \frac{1}{2} \sum_{i=1}^{N} \int_{r_{i-1}}^{r_i} (r - r_i)^2 \, r \, e^{-\frac{r^2}{2}} \, dr \tag{3.4-7}$$

Note that the energy for the normalized ($\sigma = 1$) Rayleigh process is 2,

$$\int_0^\infty r^3 e^{-r^2/2} \, dr = 2 \tag{3.4-8}$$

Evaluation of Equation (3.4-7) requires numerical techniques. The appropriate numerical integration utilized a Hermitian sixth order formula. Each region, $[r_{i-1}, r_i]$, was evaluated at 100 equidistant values. In addition to the integrand (denoted by R), the particular numerical integration requires evaluation of the first and second derivatives as well.

$$R(r) = \left(r^3 - 2r^2 \hat{r}_i + \hat{r}_i^2 r \right) e^{-r^2/2} \qquad (3.4-9)$$

$$R'(r) = \left(-r^4 + 2r^3 \hat{r}_i + \left(3 - \hat{r}_i^2 \right) r^2 - 4r \hat{r}_i + \hat{r}_i^2 \right) e^{-r^2/2}$$

$$(3.4-10)$$

$$R''(r) = \left(r^5 - 2\hat{r}_i r^4 + \left(\hat{r}_i^2 - 7 \right) r^3 + 10 \, r^2 \hat{r}_i \right.$$

$$(3.4-11)$$

$$\left. + \left(6 - 3\hat{r}_i^2 \right) r - 4\hat{r}_i \right) e^{-r^2/2}$$

Numerical integration is performed over each of the N sections and summation then performed over the N sections. The RMSE due to phase or amplitude quantization is shown in Table 3.4-1. The relative importance of phase over amplitude is effectively demonstrated by this table, particularly for coarse quantization. Ignoring amplitude completely causes 21.5 percent error of the total image energy. The single-level quantizer collapses the entire range

of the random variable into a single a priori known value. Conse-
quently, all randomness associated with that variable is removed,
thus the associated entropy is zero. Essentially, the same result is
obtained in holography (Kermish, 1970) utilizing a much more com-
plicated physical model. The phase requires 2 bits (N = 4) to main-
tain the same amount of MSE that is achieved by zero bits for ampli-
tude. Similarly, 1 bit amplitude is "worth" 3 bits of phase. Since
the majority of transform values in the experimental chapters
requires a very low degree of quantization, the quantitative results
of Table 3.4-1 are highly relevant, and demonstrative of the phase
superiority.

TABLE 3.4-1

THE RMSE INTRODUCED BY PHASE AND
AMPLITUDE QUANTIZATION

Number of Quantum Levels	RMSE Phase Quantization	RMSE Amplitude Quantization
1	2.0	0.215
2	0.73	0.042
4	0.20	0.025
8	0.05	0.011
16	0.013	0.0048
32	0.0031	0.0020

3.5 Non-Linear Effects of Phase Quantization

The general comment was made in subsection 3.5 regarding the nonlinearity of the quantizing process which is quite significant for the case of coarse quantization. The appropriate effects are structural and for them, the MSE may not be a descriptive parameter.

The importance of phase information has been emphasized. Also, the achievement of a high degree of redundancy reduction requires that most transform domain samples be quantized at few quantum levels. Therefore, it is of value to demonstrate the type of global distortion that results from quantizer nonlinearity. Specifically, coarse phase quantization will be considered.

The effect of phase quantization has been previously considered in relation to holography (Goodman and Silvesteri, 1970; and Dallas, 1971, a and b). Their analysis is applicable to image coding, with some important modifications. The primary difference is that unlike a digital image display, in holography, the final image inherently is an energy representation. Consequently, extraneous images and ghosts diminish quadratically with the number of quantum levels for holography. A similar dependence is linear for image coding, thus the distortion is more emphasized.

In the following, conventional phase definition will be utilized for the two-dimensional case. Let g and G be a Fourier transform pair:

$$\iint_{-\infty}^{\infty} g(x,y)\, e^{-2\pi j(xu + yv)}\, dxdy = G(u,v) = \left|G(u,v)\right|\, \exp j\theta(u,v)$$

$$(3.5-1)$$

The phase, θ, is linearly quantized to N levels and the inverse Fourier transform is performed. The result is denoted by $\widehat{g}(x,y)$ and it is of the following form (Dallas, 1971, a):

$$\widehat{g}(x,y) = \sum_{m=-\infty}^{\infty} \operatorname{sinc}\left\{m + \frac{1}{N}\right\} g_m(x,y) \qquad (3.5-2)$$

The sinc function can be expanded as

$$\operatorname{sinc}(m + 1/N) = \operatorname{sinc}(1/N)\,(-1)^m/(mN + 1)$$

and g_m is defined as

$$g_m(x,y) = \iint_{-\infty}^{\infty} \left|G(u,v)\right|\, \exp j(mN + 1)\,\theta(u,v)$$

$$(3.5-3)$$

$$\times \exp 2\pi j(ux + vy)\, du\, dv$$

Note that for $m = 0$, the g_m is the original image. For $m \neq 0$, g_m represents extraneous images or "ghosts."

The following additional observation can be made

a)   From Parseval theorem:

$$\int\int_{-\infty}^{\infty} \left| g_m(x,y) \right|^2 dx\ dy = \int\int_{-\infty}^{\infty} \left| g_n(x,y) \right|^2 dx\ dy \qquad (3.5\text{-}4)$$

for all integers m and n

b)   Each ghost image decreases in intensity by the factor $1/(mN + 1)$ relative to the unquantized original.

c)   In holography, as a result of the squaring operation performed by the optical system, the ghost image intensity decrease factor is $1/(mN + 1)^2$. In digital processing, this factor is $1/(mN + 1)$.

d)   The largest ghost is g_{-1}, whose relative weight is $1/(1 - N)$ with respect to g_0.

e)   One can also observe from Equation (3.5-2) that $\lim_{N \to \infty} g = g$, since

$$\text{sinc } [m] = \begin{cases} 0\ m \neq 0 \\ \\ 1\ m = 0 \end{cases} \qquad (3.5\text{-}5)$$

In digital implementation, the continuous Fourier transform, Equation (3.5-1), is replaced by its discrete equivalent, the Fourier series. The implied periodicity of the latter results in the reappearance of ghost images which have been cyclically shifted out of the basic image region.

It is possible to interpret the various ghost images, and the display of the distorted image can be quite dramatic. A computer experiment, similar to one which was holographically implemented by Dallas, was performed.

Except for the 64 x 64 element upper right sub-block, the "couple" image was zeroed out. This image was Fourier-transformed and the phase was uniformly quantized at $N = 2$, 3, 4, and 32 levels. The final images are reconstructed via the inverse transform. The result of the experiment is shown in Figure 3.5-1.

The worst case, $m = -1$, requires special attention for the two-level quantizer. Note that the weight factor for this case is identical for $m = -1$, and $m = 0$. Furthermore, from Equation (3.5-3)

$$g_{-1}(x, y) = \iint_{-\infty}^{\infty} |G(u, v)| \exp[-j\theta(u, v)] \exp 2\pi j(ux + vy) \, du \, dv$$

$$(3.5-6)$$

or, equivalently

$$\left| g_{-1}(x, y) \right| = \left| g_0(-x, -y) \right| \qquad (3.5-7)$$

By experimental construction, $g_0(-x, -y)$ does not overlap with $g_0(x, y)$, the largest ghost image is the "mirror image" of the original.
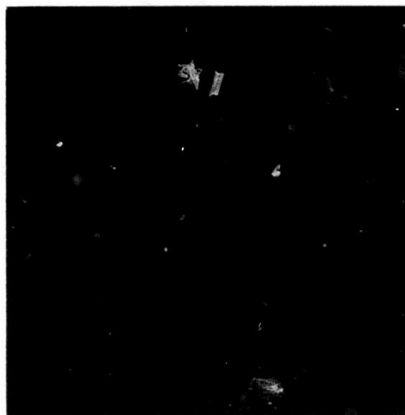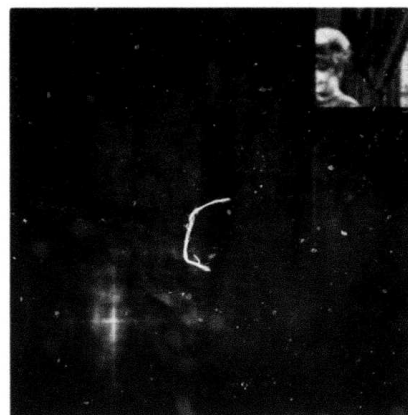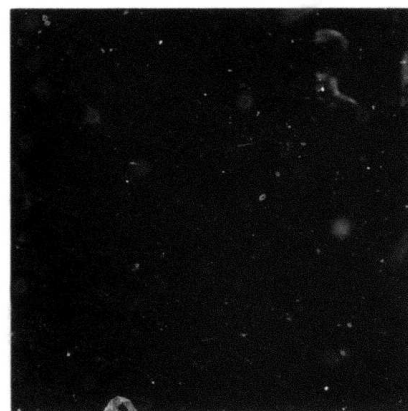
(a) Original
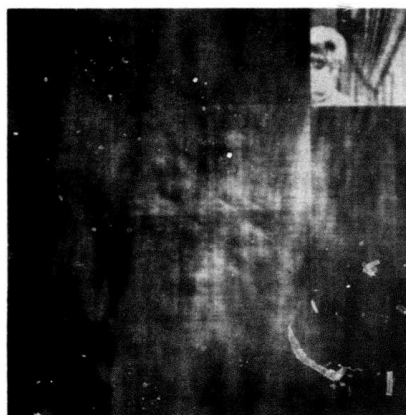
(b) 2 Level Quantizer

(c) 3 Level Quantizer

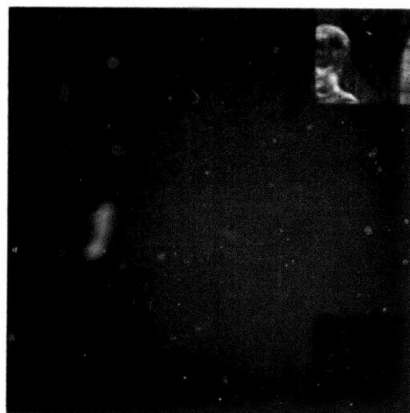(d) 4 Level Quantizer

(e) 32 Level Quantizer

Figure 3.5-1. Demonstration of Phase Quantization Effects for Fourier Domain

Although the above described experiment is rather specialized, it does emphasize the importance of the global nature of distortion introduced by the phase quantizer nonlinearity. Availability of the relevant MSE provides little if any information about the nature of the distortion.

The digital experiment was repeated for the Walsh domain and results are shown in Figure 3.5-2. The significant difference can be explained by the symmetry of the decomposition rather than by the functional properties of the eigenfunctions. Actually, the analysis related to Figure 3.5-2 is much simpler than the one associated with the Fourier case.

Consider the decomposition of $f(x_i, y_j)$ in terms of the even and odd Walsh function:

$$
\begin{aligned}
f(x_i, y_j) = \sum_{l=0}^{N/2-1} \sum_{k=0}^{n/2-1} \Big\{ &b_{cc}(k, l)\, cal(k, x_i)\, cal(l, y_j) \\
&+ b_{sc}(k, l)\, sal(k, x_i)\, cal(l, y_j) \\
&+ b_{cs}(k, l)\, cal(k, x_i)\, sal(l, y_j) \\
&+ b_{ss}(k, l)\, sal(k, x_i)\, sal(l, y_j) \Big\}
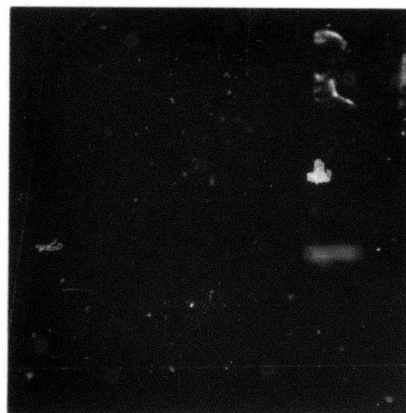\end{aligned}
$$

$$(3.5-8)$$

(a) 2 Level Quantizer
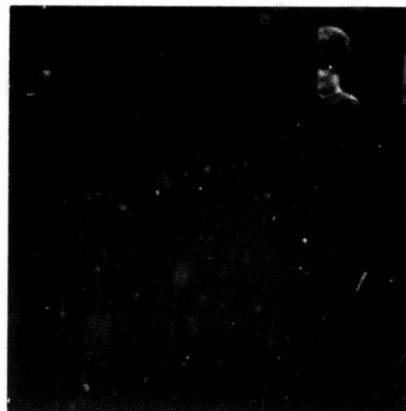
(b) 3 Level Quantizer

(c) 4 Level Quantizer

(d) 32 Level Quantizer

Figure 3.5-2. Demonstration of Phase Quantization
Effects for Walsh Domain

The conventional sequency-ordered Walsh transform will yield the "b" matrix for the two-dimensional case. Consider the following "unconventional" phase definition such that

$$f(x_i, y_j) = \sum_{l=0}^{N/2-1} \sum_{k=0}^{N/2-1} B_c(k, l) \cos \theta_1(k, l) \, cal(k, x_i) \, cal(l, y_j)$$

$$+ B_c(k, l) \sin \theta_1(k, l) \, sal(k, x_i) \, cal(l, y_j)$$

$$+ B_s(k, l) \cos \theta_2(k, l) \, cal(k, x_i) \, sal(l, y_j)$$

$$+ B_s(k, l) \sin \theta_2(k, l) \, sal(k, x_i) \, sal(l, y_j) \qquad (3.5-9)$$

where

$$B_c^2 = b_{cc}^2 + b_{sc}^2$$

$$B_s^2 = b_{cs}^2 + b_{ss}^2$$

$$\theta_1 = \tan^{-1}(b_{sc}/b_{cc})$$

$$\theta_2 = \tan^{-1}(b_{ss}/b_{cs})$$

For the particular original of Figure 3.5-1, the coefficients are equal, $b_{cc}(k, 1) = b_{ss}(k, 1) = b_{sc}(k, 1) = b_{cs}(k, 1) = b(k, 1)$. This can be shown by letting $f(x_i, y_j) \neq 0$ in Equation(3.5-18). Because of the symmetry of the image, it follows that $f(-x_i, y_j) = f(x_i, -y_j)$ $f(-x_i, -y_j) = 0$. Simple algebraic manipulation of Equation (3.5-18) will yield the equality of the coefficients. Consequently, $\theta_1 = \theta_2 = \pi/4$.

If $f(x_i, y_j)$ is the image in Figure 3.5-1a Equation (3.5-8) becomes

$$f(x_i, y_j)' = \sum_{1=0}^{N/2-1} \sum_{k=0}^{N/2-1} b(k, 1) \Big\{ cal(k, x_i)\, cal(1, y_j) \\ + sal(k, x_i)\, cal(1, y_j) \\ + cal(k, x_i)\, sal(1, y_j) \\ + sal(k, x_i)\, sal(1, y_j) \Big\}$$

$$(3.5-10)$$

Consider the application of a two-level uniform phase quantizer; Equation (3.5-10) will become

$$f(x_i, y_j) = \sum_{1=0}^{N/2-1} \sum_{k=0}^{N/2-1} \frac{2}{\sqrt{2}} b(k,1)\Big\{sal(k,x_i)\,cal(1,y_j)$$

$$+ sal(k,x_i)\,sal(1,y_i)\Big\}$$

$$(3.5-11)$$

The result for the three-level quantizer is

$$f(x_i, y_j) = \sum_{1=0}^{N/2-1} \sum_{k=0}^{N/2-1} \frac{2}{\sqrt{2}} b(k,1)\Big\{cal(k,x_i)\,cal(1,y_j)$$

$$+ cal(k,x_i)\,sal(1,y_j)\Big\}$$

$$(3.5-12)$$

For the four-level quantizer, the quantized result is identical to the original (unquantized). The symmetry of the quantized images in Figure 3.5-2 is equivalent to the symmetry expressed by the related Equations (3.5-21) and (3.5-22).

The Walsh domain phase quantization experiment provides another indication regarding the nonlinear nature of the quantizer.

The phase definitions θ_1 and θ_2 may appear artificial, however, it is convenient in the sense that they are defined on adjacent transform pairs in the conventionally ordered two-dimensional Walsh transform.

3.6   The α Processor

In the context of the phase superiority vs amplitude, the per-formance of the so-called α processor is not unexpected.

Consider the image Fourier transform for the two-dimensional case

$$\widetilde{I}(u, v) = \int\!\!\!\int_{-\infty}^{\infty} I(x, y)\, e^{-2\pi j(xu + yv)}\, dx\, dy$$

$$= \left| \widetilde{I}(u, v) \right|\, e^{-j\theta(u, v)} \tag{3.6-1}$$

The α processor is defined as the nonlinear operator, T_{α} which raises the transform amplitude to the power α:

$$T_{\alpha}\{\widetilde{I}(u, v)\} = \left| \widetilde{I}(u, v) \right|^{\alpha}\, e^{-j\theta(u, v)} \tag{3.6-2}$$

Consider the effect for $\alpha \epsilon [0, 1]$. One can explicitly designate the transform amplitude, $I(u, v)$ by two terms where $R(u, v)$ is the image power spectral density and $r(u, v)$ the amplitude fluctuation around the power spectral density:

$$\left| \widetilde{I}(u, v) \right| = R(u, v) + r(u, v) \tag{3.6-3}$$

therefore,

$$H(u, v)\, \widetilde{I}(u, v) = \{H(u, v)\, R(u, v) + H(u, v)\, r(u, v)\} e^{-j\theta(u, v)}$$

$$\tag{3.6-4}$$

Here H(u, v) is a linear filter. Consider the application of T_α $\alpha \epsilon [0, 1]$:

$$T_\alpha \widetilde{I}(u, v) = \left| I(u, v) \right|^\alpha e^{-j\theta(u, v)}$$

$$= \left| R(u, v) + r(u, v) \right|^\alpha e^{-j\theta(u, v)}$$

$$\simeq R^\alpha + \alpha R^{\alpha - 1} r \tag{3.6-5}$$

The ratio of the amplitude fluctuation and the power spectral density has decreased in Equation (3.6-5) from $r(u, v)/R(u, v)$ to $\alpha r(u, v)/R(u, v)$. One guarded observation is that amplitude entropy has decreased by an amount related to $(1 - \alpha)$.

For $\alpha = 0$, image transform amplitudes are identically unity. Consequently, the image in this case became a white process, since image power spectral density is also a constant. The $\alpha = 0$ case demonstrates two interesting image properties. First, under conventional ergodic assumptions the image becomes uncorrelated. Yet, visual inspection of the appropriate images indicates (Figure 3.6-1) that basic image features have not changed. The $\alpha = 0$ filter drastically changed image statistics, yet the apparent visual image structure remained relatively unaltered.

3.7 Phase-Only Image (Polynomial Magnitude Fit)

The "α processor" has decreased the amplitude entropy in the transform domain, however, it also changed the image power spectral density. It is important to separate the two effects. An approximate

(a) Walsh Transform
$\alpha = 0.5$

(b) Walsh Transform
$\alpha = 0$

(c) Fourier Transform
$\alpha = 0.5$

(d) Fourier Transform
$\alpha = 0$

Figure 3.6-1. Demonstration of the α Processor

This page is reproduced at the back of the report by a different reproduction method to provide better detail.

linear inverse filter to T_α could restore the power spectral density to its original form. A more straightforward technique is to fit a particular type of surface to the image transform amplitudes. This second approach is considered in this section.

Consider the two dimensional transform domain of an image, $\tilde{I}(u, v)$. It is not necessary to specify the particular transform. The discretized version of $\tilde{I}$ will be used, such that transform parameters u and v are integers.

The image transform amplitudes will be least square fitted by a two-dimensional surface, $Z(u, v)$ of the following form:

$$Z(u, v) = R(u, v)\left\{a_{00} + a_{10}\, u + a_{01}\, v + a_{20}\, u^2 + \dots + a_{0N}\, v^N\right\}$$

$$(3.7-1)$$

or in a more compact notation:

$$Z(u, v) = R(u, v) \sum_{j=0}^{N-1} \sum_{i=0}^{N} a_{ij}\, u^i\, v^j \qquad (3.7-2)$$

The weight function $R(u, v)$ is specified in advance and the coefficients a_{ij}'s are the unknowns to be determined. For a given N, the number of coefficients is $1/2\,(N + 1)\,(N + 2)$.

The mathematical objective is to minimize the mean-square deviation between $\left| \tilde{I}(u, v) \right|$ and $Z(u, v)$, that is

$$\sum_{u} \sum_{v} \left\{ R(u,v) \sum_{j=0}^{N-1} \sum_{i=0}^{N} a_{ij} u^i v^j - \left| \tilde{I}(u,v) \right| \right\}^2 = \text{minimum}$$

$$(3.7-3)$$

The minimization is accomplished by differentiating (3.7-3) with respect to a_{kl} for $k = 1, \ldots N$; $l = 1, \ldots N - k$ and solving the $1/2 (N+1)(N+2)$ linear equations.

$$\sum_{u} \sum_{v} \left\{ R^2(u,v) \left[\sum_{j=0}^{N-1} \sum_{i=0}^{N} a_{ij} u^i v^j \right] u^k v^l \right\}$$

$$= \sum_{u} \sum_{v} R(u,v) \left| \tilde{I}(u,v) \right| u^k v^l \qquad (3.7-4)$$

Equation (3.7-4) can be rewritten in the following matrix notation:

$$\sum_u \sum_v \left\{ R^2(u,v) \begin{bmatrix} 1 \\ u \\ & v \\ & & uv \\ & & & \cdot \\ & & & & \cdot \\ & & & & & \cdot \\ & & & & & & v^N \end{bmatrix} \begin{bmatrix} 1 & u & v & uv & .. & v^N \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ 1 & u & v & uv & .. & v^N \end{bmatrix} \right\} \begin{bmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \\ \\ a_{0N} \end{bmatrix}$$

$$= \sum_u \sum_v \begin{bmatrix} R(u,v) \left| \widetilde{I}(u,v) \right| \\ R(u,v) \left| \widetilde{I}(u,v) \right| u \\ R(u,v) \left| \widetilde{I}(u,v) \right| v \\ R(u,v) \left| \widetilde{I}(u,v) \right| v^N \end{bmatrix} \qquad (3.7\text{-}5)$$

Equation (3.7-5) is in the form of a conventional linear matrix equation with the column matrix of the a_{ij}'s being the unknown. For a given image transform and a specified weight function, Equation (3.7-5) may be solved by many conventional techniques for the solution of systems of linear equations (Blum, 1972).

The actual least square fit is dependent on the choice of the weight function, $R(u,v)$. Note also that $Z^2(u,v)$ is an estimate of the image power spectral density. Any a priori image information should be incorporated into $R(u,v)$.

Under the conventional separable Markov model, the image correlation function is of the form $e^{-\alpha(|x| + |y|)}$ and the appropriate power spectral density is of the form $\sim[(u^2 + \alpha^2)(v^2 + \alpha^2)]^{-1}$. A reasonable choice of $R(u, v)$ can therefore be picked as

$$R(u, v) = [(u^2 + \alpha^2)(v^2 + \alpha^2)]^{-1/2} \qquad (3.7\text{-}6)$$

For the adjacent element correlation of 0.95 the value of α is ln 0.95.

The utilization of Equation (3.7-6) for the least square fit problem indicates how good (or bad) the Markov model is. If the Markov model perfectly represented the image statistics, except for the A_{00} term, all other coefficients would be zero. The ratios of the appropriate coefficients (e.g., A_{ij}/A_{00}, $i + j > 0$) provides quantitative information on the deviation between the actual power spectral density and the one predicted by the Markov model.

The replacement of the individual amplitude values by the appropriate related power spectral density values provides an important demonstration of the phase superiority. The L^2 image values for an L X L image are represented by $1/2\ L^2$ amplitude and $1/2\ L^2$ phase values. The power spectral density surface is prescribed by a negligible (relative to $1/2\ L^2$) number of coefficients. The availability of the least square fitted surface permits the replacement of $1/2\ L^2$ values, in effect, by a few parameters.

Equation (3.7-5) was implemented numerically utilizing the Fourier domain. The Markov model was used for the weight function

with 0.95 as the element-to-element vertical and horizontal corre-
lation. The highest degree of the two-dimensional polynomial was 2
through 5 for the five cases considered. The respective number of
terms in the polynomial ranged from 6 through 21. Table 3.7-1
shows the various cases.

Table 3.7-1

Degree and Number of Terms in the
Surface Fitting Polynomials

N	$(1/2)(N + 1)(N + 2)$
1	3
2	6
3	10
4	15
5	21

The images generated by the above-outlined procedure have
good visual correlation with the original (Figure 3.7-1). The high
spatial frequency details are completely preserved. Not unex-
pectedly, the basic apparent distortions are in the very low-frequency
region. Generally, it is the low-frequency region which does not
lend itself to good statistical characterization. The reason is that
the low-frequency amplitudes can be recovered from a very coarsely
sampled image, thus the law of large numbers which is always
implied in an ergodic approximation, does not apply. On the other

(a) Original

(b) 21 Term Expansion

(c) 15 Term Expansion

(d) 10 Term Expansion

(e) 6 Term Expansion

Figure 3. 7-1. Demonstration of Amplitude Polynomial Fit

hand, one may be too generous in allocating extra bits for the very low-frequency region since the impact on the overall bit rate will be negligible.

The pictorial representation of the actual polynomial surfaces is shown in Figure 3.7-2 while the calculated coefficients are given in Table 3.7-2. The large value for coefficients other than the a_{00} term indicates that the exponential Markov correlation model requires higher order corrections.

The amplitude surface fitting procedure could be utilized on the development of an actual transform coding algorithm, however, it was abandoned in favor of the recursive approach which is the topic of Chapters 4 through 6. The solution of Equation (3.7-5) and the recalculation of the amplitude surface is likely to generate such additional computation load in addition to the actual transform algorithm, that any practical implementation would be prohibitive. For the fifth degree polynomial approximation a 21st order matrix equation must be solved. Each surface element recalculation requires in excess of 21 addition and multiplication operations. The latter operations amount to a higher number of arithmetic steps than required by the full size Fourier transform. In addition, both the solution of the matrix equation and the reconstruction of the surface are somewhat ill-conditioned. The numerical implementations of this section were done on a 60-bit wordlength computer. It is
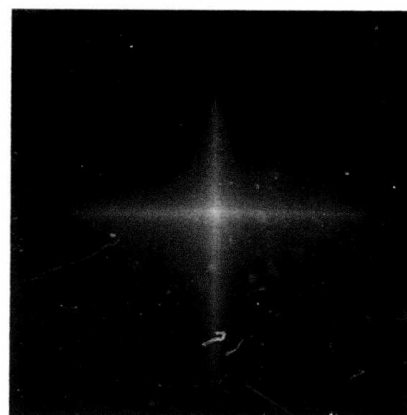
(a) Original
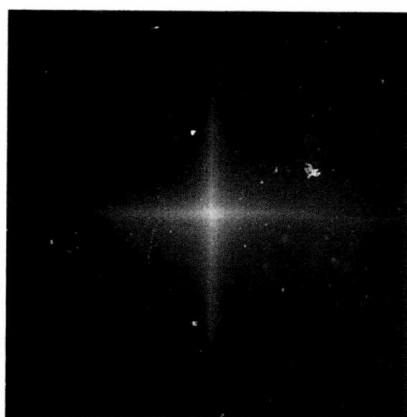
(b) 21 Term Expansion

(c) 15 Term Expansion

(d) 10 Term Expansion

(e) 6 Term Expansion

Figure 3. 7-2.   Fourier Domain Display of Polynomial
Fitted Amplitudes

anticipated that the round-off errors might not be negligible had the same calculations, particularly the matrix equation, been performed on a computer with shorter word length. In which case, the requirement for double precision would further increase the computational load of the coding-decoding procedures.

Table 3. 7-2

Calculated Coefficients for the "Girl" Image

Coefficients	Number of Coefficients			
	6	10	15	21
a_{00}	0.617	0.617	0.624	0.633
a_{10}	0.069	0.130	0.119	0.158
a_{01}	1.22	1.37	0.871	0.227
a_{20}	-3.22	- 0.695	- 1.14	- 1.93
a_{11}	-0.075	- 0.298	0.374	1.96
a_{02}	-1.57	- 2.82	1.04	8.29
a_{30}		- 0.161	- 0.156	- 0.660
a_{21}		10.1	18.7	43.4
a_{12}		0.327	- 2.70	-22.1
a_{03}		1.18	- 7.25	-35.2
a_{40}			0.423	1.57
a_{31}			- 0.226	1.79
a_{22}			-15.5	-44.0
a_{13}			2.90	49.8
a_{04}			5.37	46.0
a_{50}				0.524
a_{41}				-31.0
a_{32}				- 2.94
a_{23}				26.7
a_{14}				-30.0
a_{05}				-19.7

4. EXPERIMENTAL RESULTS I (MONOCHROME)

The concepts developed in the preceding chapters have been implemented. Computer algorithms have been developed for the coding and decoding of various images. This chapter considers the algorithm for monochrome images.

For practical reasons, the coding algorithms only included digital input and output. For the monochrome image coding examples, the input is a square image sampled over a 256 x 256 grid. Each sample is linearly quantized to 256 levels.

The significant achievement of the adaptive phase coding process discussed here is that the transmitter is slaved to the receiver without any overhead information. Yet complete adaptivity is possible, as well as arbitrary sample reduction. The drawback of adaptive procedures is the requirement for large buffers. This requirement is unavoidable but it is not likely to be important in the environment of computer-to-computer communication. In this case, the undecoded images can easily be stored, for example, on magnetic tapes.

The most demanding computational step is the large size, 256 by 256, image transforms. It is interesting to note that the computational complexity, that is the number of arithmetic operations, increase rather slowly from the case when the sub-block transforms are replaced by one single large transform. For example, the ratio of the number of operations for the entire image transform (256 x 256) vs 16 x 16 sub-blocks is log 256/log 16 = 2. A factor of two increase in arithmetic complexity is not too extreme in computer implementation.

The 256 x 256 size is generally too large to permit the two-dimensional transformation entirely in core. The cost of the additional I/O operations should also be included.

4.1 Description of the Algorithm

The importance of image representation by amplitude and phase was demonstrated in Chapter 3. In particular, the phase superiority was established. The coding algorithm should incorporate these important properties of the transform domain.

The following assumptions are made: (1) the transform values are uncorrelated and normally distributed, (2) the power spectral density, equivalent to the sample variances, is a smooth surface. It is significant to note that these assumptions are, in fact, related. It can be shown that the Fourier transform will produce uncorrelated samples under the assumption of smooth power spectral density (Papoulis, 1965; Chapter 13).

The basic two assumptions lead to the equivalent amplitude and phase representation. Furthermore, the amplitude is Rayleigh and the phase is uniformly distributed. Specifically, an N x N image is decomposed into $1/2\ N^2$ amplitude and $1/2\ N^2$ phase terms, which are, by assumption, mutually independent. Two separate coding schemes were developed depending on the transform symmetry. One coder utilizes the complex Fourier transform. The conventional odd and even function decomposition into amplitude and phase is used by the second coding scheme for which the Walsh transform is

used. The two types of representation are related by a simple mapping, thus either coding scheme is sufficient for both Fourier and Walsh transformations. The schematics of the coding-decoding process are shown in Figure 4.1-1.

Detailed descriptions of the algorithms for both the Fourier and Walsh transform are provided in this section.

The 256 x 256 image is Fourier-transformed. The conventional representations of the Fourier domain are shown in Figures 4.1-2 and 4.1-3. The arrows indicate increasing harmonics in the horizontal and vertical directions. The number pairs in parentheses indicate the ordering of the amplitudes (or phases) according to harmonics. The discrete fast numerical transform yields Figure 4.1-2. The more familiar diffraction pattern is shown in Figure 4.1-3. By interc'.anging the two halves of the pattern either representation can be easily mapped into the other one.

An ordering must be established which specifies the sequence for the Fourier domain qua. ization. The rows are indexed according to the natural ordering. Referring to Figure 4.1-3, the first row is the top of the pattern and then the coder proceeds downward. The significant practical advantage of this scheme is that the computer algorithm will not require a large memory block. It is not required to store more than a small fraction of the discrete transform in memory; this, however, depends on the complexity of the predicting algorithm.
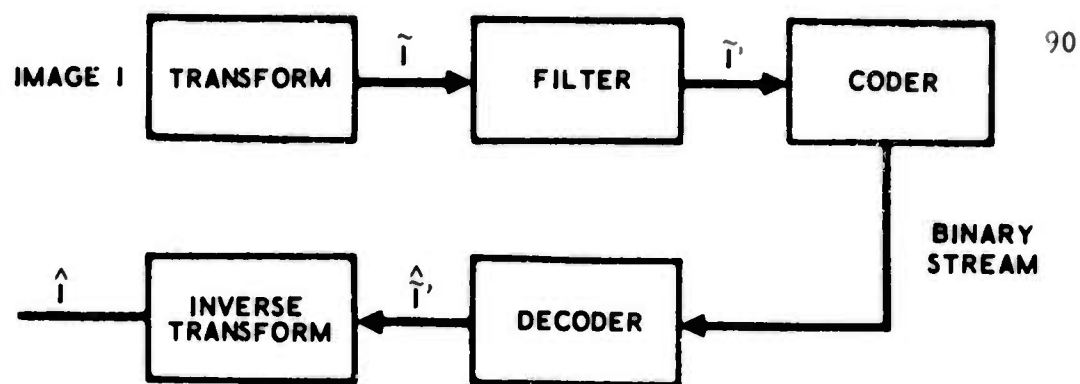
Figure 4.1-1. Schematic of the Coding-Decoding Procedure
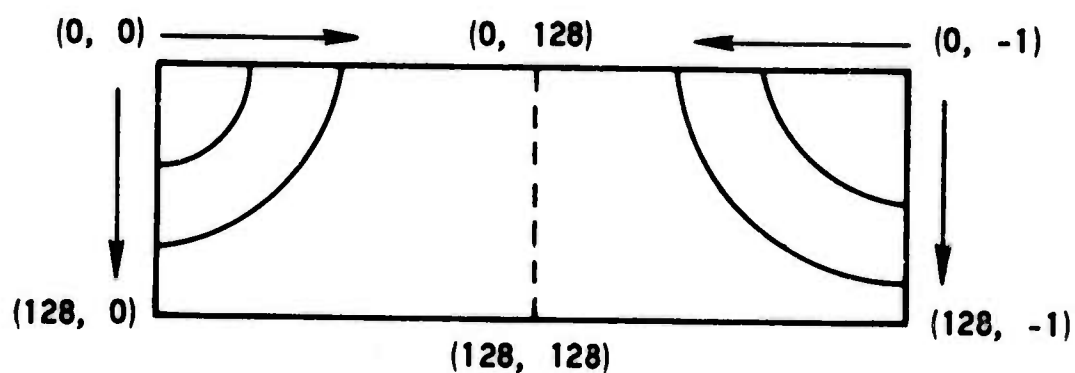
Figure 4.1-2. Conventional Fourier Domain Representation I
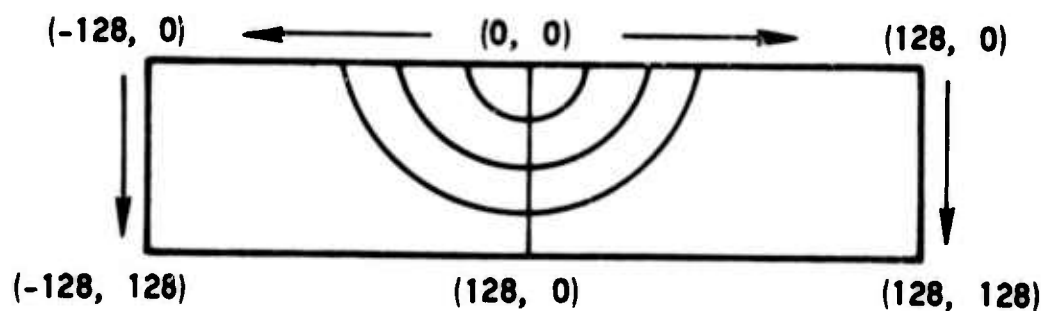
Figure 4.1-3. Conventional Fourier Domain Representation II
(Note: Columns 128 and -128 are identical)

Within each row, the coder starts with the lowest horizontal harmonic, then it proceeds to the right (refer again to Figure 4.1-3) following which it repeats the process moving to the left from the center.

The code words are generated by the quantization of the amplitude and phase values. The phase values are uniformly quantized. The amplitude is companded and then processed by a uniform quantizer. The number of quantum levels is set in linear proportion to the variance of the transform samples. The number of quantum levels for the phase is twice as high as that for the amplitudes, when this number is four or larger. For the two-level amplitude quantization, eight phase quantum levels are specified. The transform domain variance is estimated from the previously quantized amplitude values. Clearly, the estimate based on the amplitudes prior to quantization would be preferable, however, it would lead to an undecodable process. The decoder will also perform the estimation process and it only has access to the previously quantized amplitudes.

Estimation of the variance of the next amplitude to be quantized follows a rather simple rule. The density function for the Rayleigh distribution is given by

$$P(x) = \frac{x}{2\pi\sigma^2} \, e^{-\frac{x^2}{2\sigma^2}} \quad x \geq 0 \qquad (4.1\text{-}1)$$

$$= 0 \qquad \text{otherwise}$$

The first two moments are

$$E(x) = \sqrt{\frac{\pi}{2}} \ \sigma \qquad\qquad (4.1-2)$$

$$E\left\{x^2 - E(x)\right\} = VAR(x) = \left(2 - \frac{\pi}{2}\right)\sigma^2 \qquad (4.1-3)$$

For each transform amplitude, the compander needs σ, and the number of quantum levels is determined by the variance. Equations (4.1-2) and (4.1-3) can be rewritten in a more useful form as

$$\sigma = \sqrt{\frac{2}{\pi}} \ E(x) \qquad\qquad (4.1-4)$$

$$VAR(x) = \left(\frac{4}{\pi} - 1\right) E^2(x) \qquad\qquad (4.1-5)$$

Equations (4.1-4) and (4.1-5) indicate that the estimate of the average amplitude also specifies the standard deviation and the variance. The amplitude estimate is determined by averaging the previously quantized amplitudes in a neighborhood surrounding the estimate. This neighborhood is determined by the ordering of the transform domain.

The d.c. value is transmitted without requantization. The estimate of this term is, therefore, perfect. The estimate of the next value is also the d.c. value. This term will be quantized and the reconstructed value is available for the estimate of the next amplitude. The estimate of the third value is the arithmetic mean of the d.c. term and the first quantized harmonic. For all other

values on the first row, the amplitude estimate is the average of the three previously reconstructed terms.

The estimation of the amplitudes on the zero column (the column containing the d.c. term) is the exact symmetrical equivalent of the first row. All other estimates are generated from four previously quantized values by simple averaging. These samples are three values from the previous row and the just previously quantized amplitude on the same row. Equations (4.1-6) through (4.1-15) are mathematical forms of these sample estimates. The subscripts refer to the horizontal and vertical harmonic ordering of Figure 4.1-3.

$$\hat{x}_{0,0} = x_{0,0} \tag{4.1-6}$$

$$\hat{x}_{0,1} = \hat{x}_{0,0} \tag{4.1-7}$$

$$\hat{x}_{0,2} = (\hat{x}_{0,0} + \hat{x}_{0,1})/2 \tag{4.1-8}$$

$$\hat{x}_{0,j} = (x_{0,j-1} + \hat{x}_{0,j-2} + \hat{x}_{0,j-3})/3, \ j > 2 \tag{4.1-9}$$

$$x_{0,j} = (\hat{x}_{0,j+1} + \hat{x}_{0,j+2} + \hat{x}_{0,j+3})/3, \ j < 0 \tag{4.1-10}$$

$$\hat{x}_{1,0} = \hat{x}_{0,0} \tag{4.1-11}$$

$$\hat{x}_{2,0} = (\hat{x}_{1,0} + \hat{x}_{0,0})/2 \tag{4.1-12}$$

$$\hat{x}_{1,0} = (\hat{x}_{i-1,0} + \hat{x}_{i-2,0} + \hat{x}_{i-3,0})/3, \ i > 2 \tag{4.1-13}$$

$$\hat{x}_{i,j} = (\hat{x}_{i,j-1} + \hat{x}_{i-1,j-1} + \hat{x}_{i-1,j} + \hat{x}_{i-1,j+1})/4, \ i > 0, \ j > 0$$

$$(4.1-14)$$

$$\hat{x}_{i,j} = (\hat{x}_{i,j+1} + \hat{x}_{i-1,j-1} + \hat{x}_{i-1,j} + \hat{x}_{i-1,j+1})/4, \ i > 0, \ j < 0$$

$$(4.1-15)$$

The estimation of the zero row [Equations (4.1-6) through (4.1-10)] and the zero column [Equations (4.1-11) through (4.1-13)] are separated from the general form of estimation [Equation (4.1-14) for the right and Equation (4.1-15) for the left half of the Fourier plane]. The zero row and column usually have a higher degree of energy concentration than their immediate neighborhood, due to windowing, and thus require special consideration.

The mapping utilized by the compander is the distribution function associated with the appropriate probability density function. It is given by the following expression for the Rayleigh distribution:

$$y = F(x) = 1 - e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}, \ x \epsilon [0, \infty] \qquad (4.1-16)$$

Its inverse is

$$x = F^{-1}(y) = \sigma \sqrt{-2 \ln(1 - y)}, \ y \epsilon [0, 1] \qquad (4.1-17)$$

In terms of the previous equations, the coding-decoding process may be explicitly specified (see Figures 4.1-4 through 4.1-6).
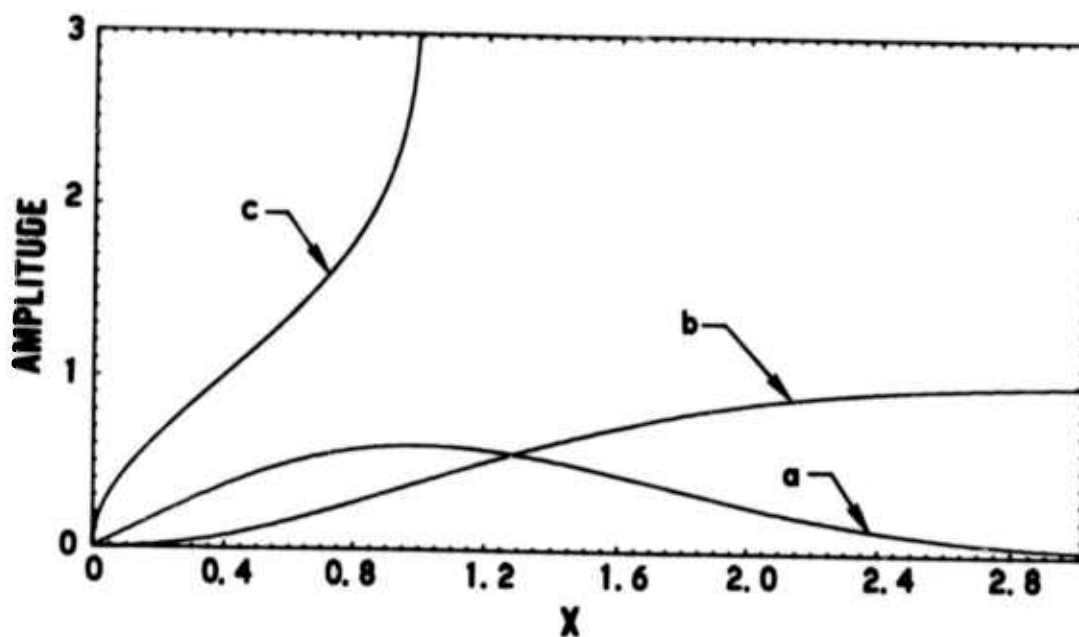
Figure 4.1-4. Various Functions Associated with Companding the Unit Variance Rayleigh Process. (a) Density Function (Equation 4.1-1), (b) Companding Transform (Equation 4.1-16), (c) Inverse Mapping (Equation 4.1-17)
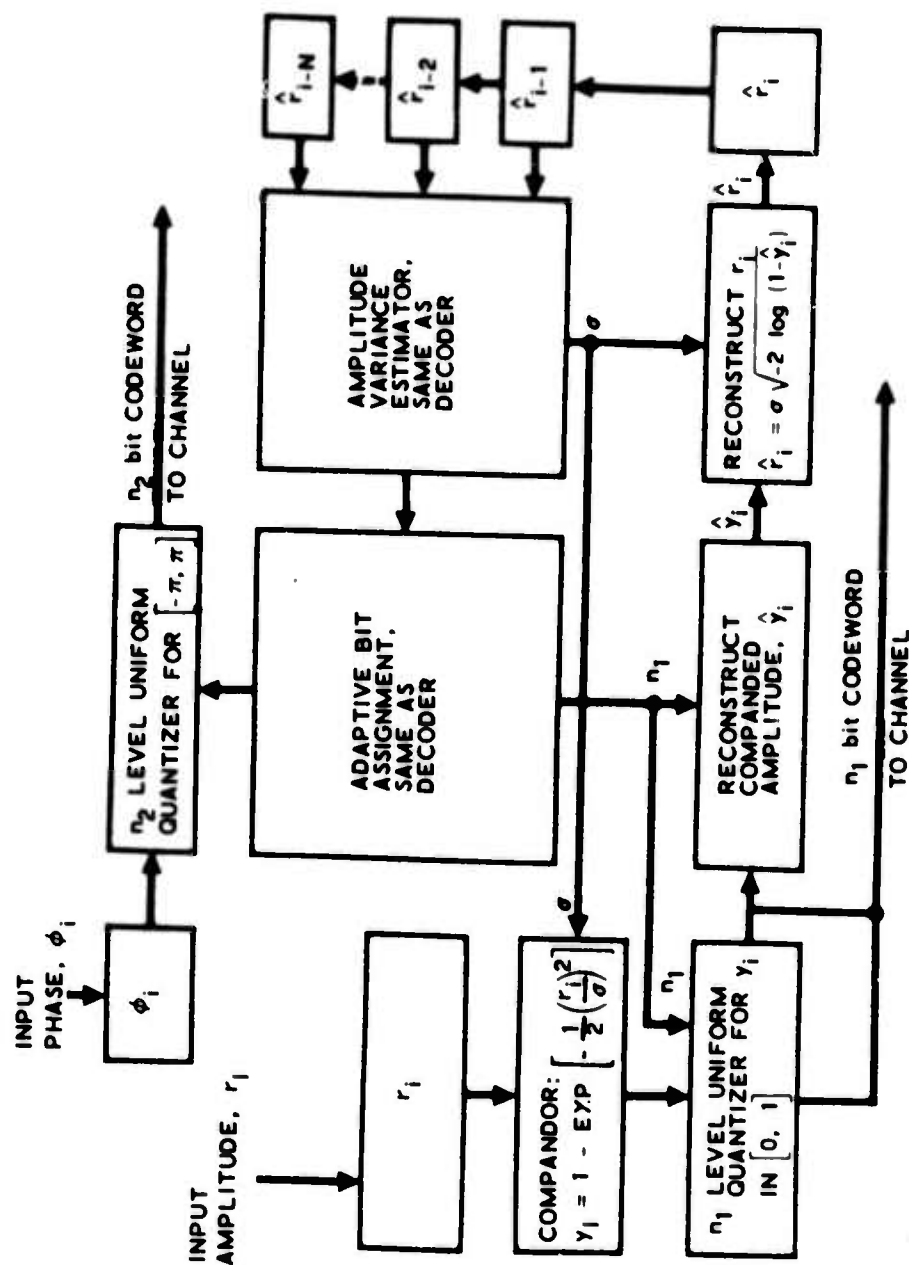
Figure 4.1-5. Adaptive Coding Algorithm
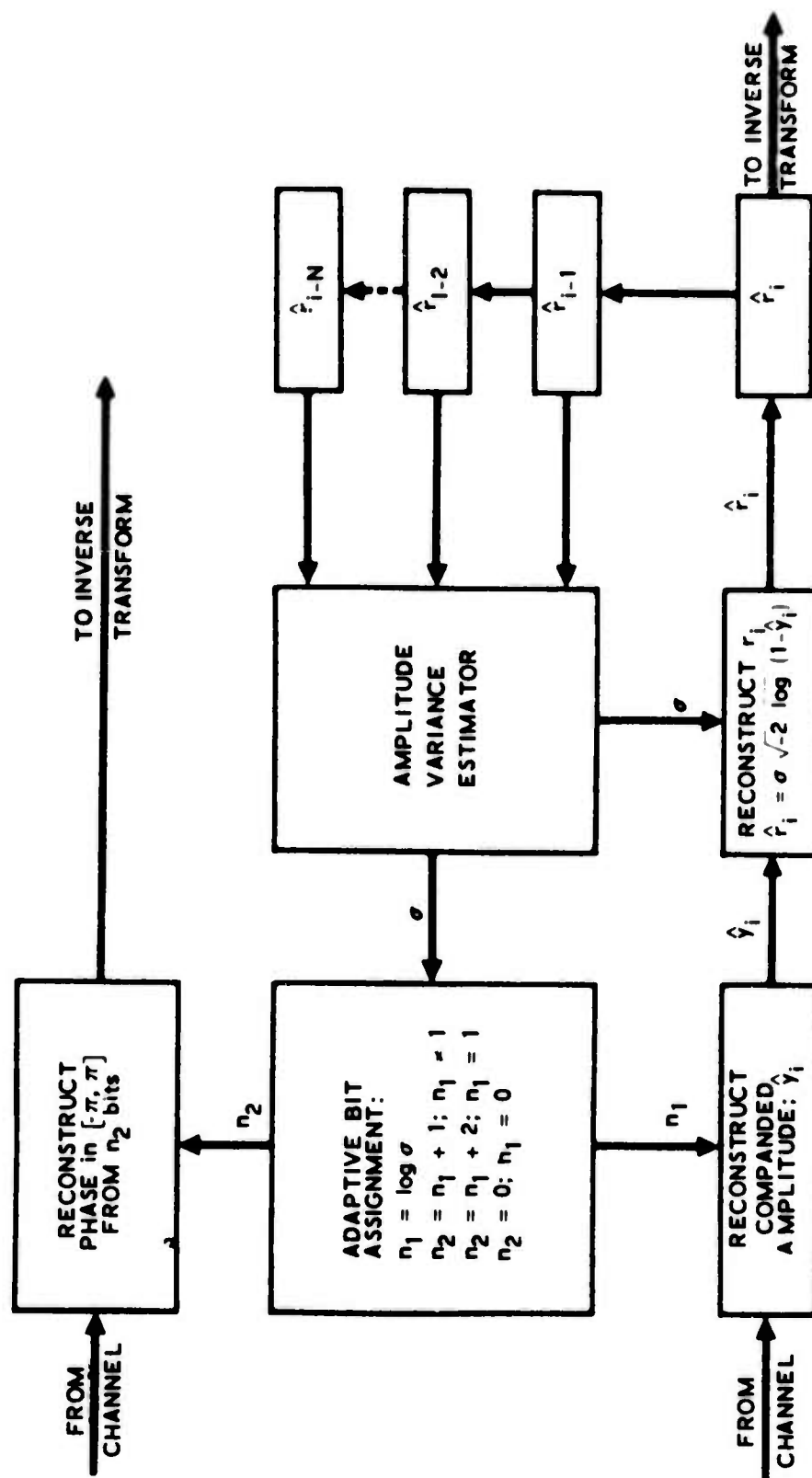
Figure 4.1-6. Adaptive Decoding Algorithm

Coding Steps

1) Transmit d. c. phase and amplitude "perfectly"

2) Estimate current amplitude from those previously quantized, utilizing one of the set of Equations (4. 1-6) through (4. 1-15)

3) Determine variance of Rayleigh distribution from Equations (4. 1-4) and (4. 1-5) by letting $E(x) \triangleq x_{i, j}$

4) Compand amplitude through Equation (4. 1-16)

5) Specify the number of quantum levels, 2^N according to the amplitude variance

6) Quantize companded amplitude and phase by uniform quantizer and transmit the appropriate code word. (Its length is $2N + 1$ bits if more than 2-level amplitude quantizer is used, otherwise it is 4.)

7) Utilizing Equation (4. 1-17), determine the actual reconstructed amplitude and save for further estimation

8) Unless the entire transform plane is quantized proceed to Step 2 for the next amplitude and phase value processing.

Decoding Steps

1) Receive exact d. c. phase and amplitude

2) Estimate current variance of Rayleigh distribution from Equations (4. 1-4) and (4. 1-5) by letting $E(x) = \hat{x}_{i, j}$; $\hat{x}_{i, j}$ is determined via the estimator Equations (4. 1-6) through (4. 1-15)

3)   Determine code word length from the amplitude variance

4)   Reconstruct inputs to the uniform quantizer (this is the companded amplitude and phase)

5)   Reconstruct amplitude utilizing Equation (4.1-17)

6)   Unless the entire transform plane is decoded, proceed to Step 2 for the next amplitude and phase decoding.

Several important observations should be made at this point. The coding-decoding process is clearly decodable. The decoding must be done in the same order in which the encoder operated. In other words, selected decoding of individual code words or sequence of code words is not possible. The code words are clearly of the variable-length type. The set of binary digits which represents the entire coding process does not possess any particular algebraic properties. It should be pointed out that although the coding process is decodable the actual binary sequence is not decodable according to the conventional definition of algebraic decodability. The quantization of the Rayleigh process can effectively be demonstrated via input-output diagrams as shown in Figures 4.1-7 through 4.1-10.

Since the decoding process is recursive, the errors made in the decoding process can be catastrophic. A catastrophic error will propagate throughout the decoding process, thus, all decoded values will be in error past the one in the sequence where the first error occurred. The primary source of error is channel noise which will be considered in Section 4.4. A catastrophic error will occur when
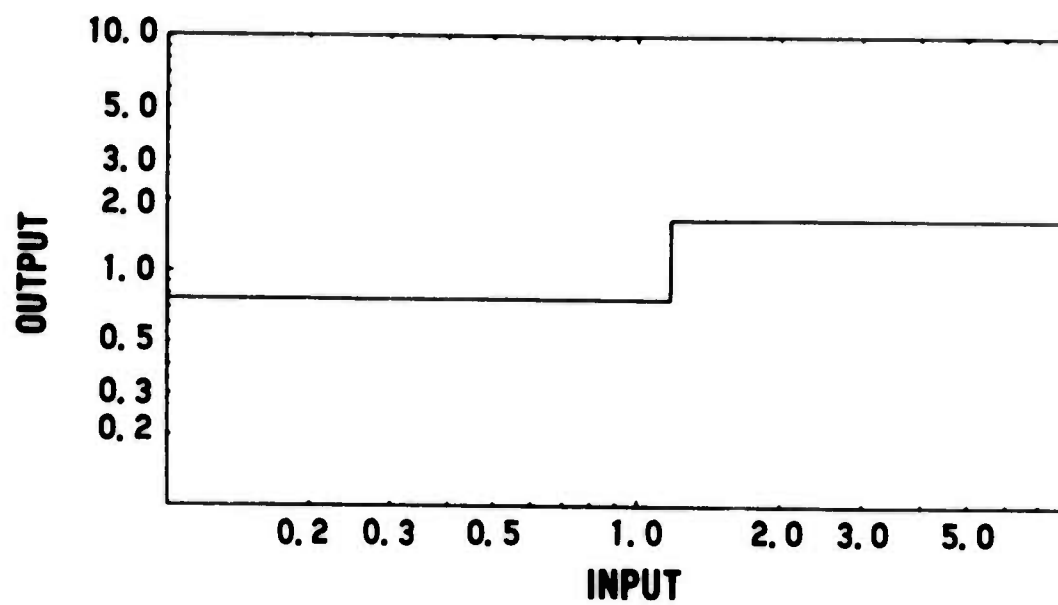
Figure 4. 1-7.   Two-Level Quantizer
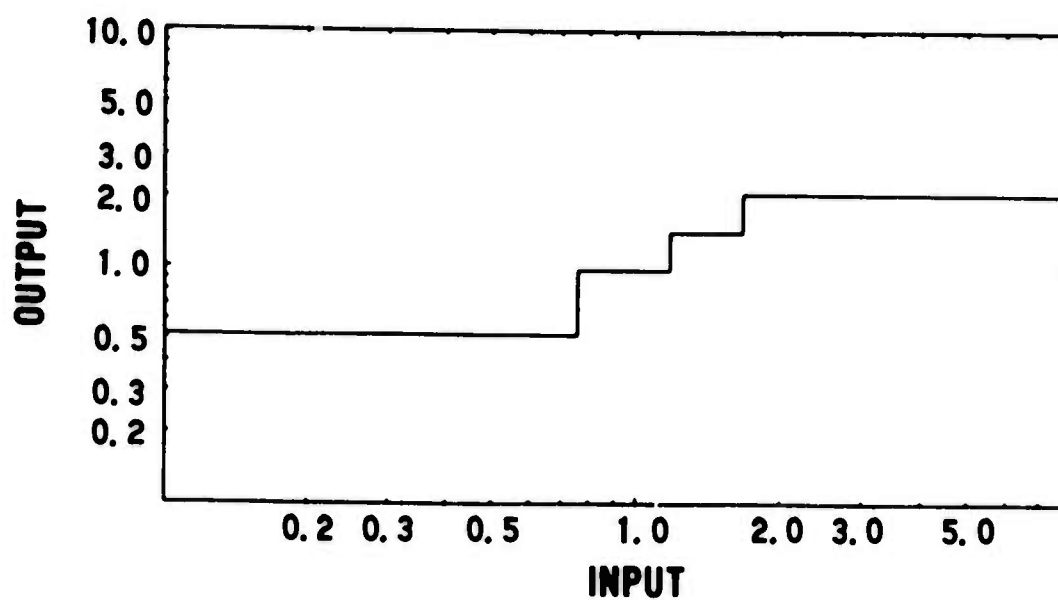
Figure 4. 1-8.   Four-Level Quantizer
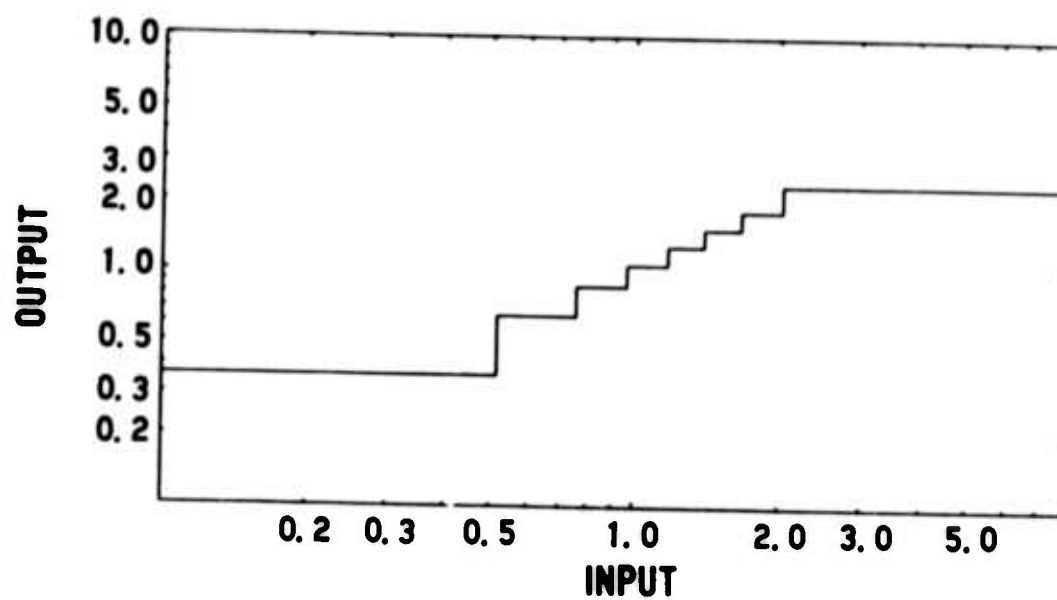
Figure 4.1-9.   Eight-Level Quantizer
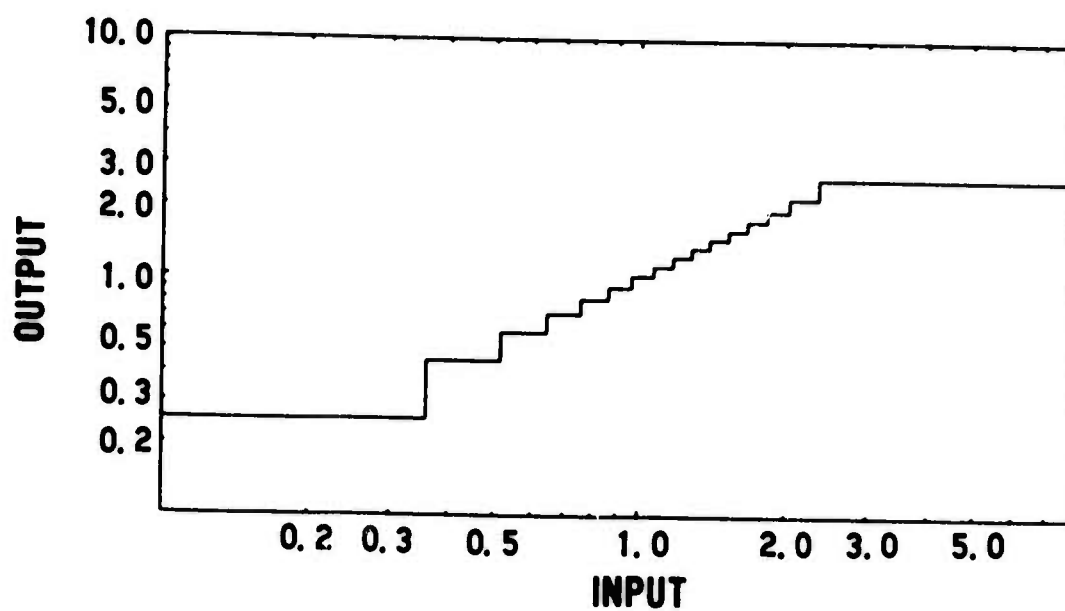
Figure 4.1-10.   Sixteen-Level Quantizer

the estimated variance is sufficiently incorrect to yield an incorrect bit assignment. The result is loss of synchronization.

The bit assignment is based on Equation (4.1-5) and it is simply of the form

$$N = [\alpha \log_2 (VAR(x))] \qquad (4.1-18)$$

As previously indicated, the amplitude and phase, in general, are specified by $2N + 1$ or $2N + 2$ bits. The α is the proportionality constant; the brackets [ ] specify the largest integer whose value does not exceed the value within the brackets. Both the encoding and decoding algorithms include a large number of arithmetic operations; specifically, Equations (4.1-4) through (4.1-17) are utilized before Equation (4.1-18) can be applied. In order to assure that the result of Equation (4.1-18) is identical for both the encoding and decoding processes, it is important that the sequence and accuracy of the arithmetic operations be the same. The correct sequence is achieved by proper programming. The accuracy consideration is much more involved. Clearly, if the coding and decoding algorithms are implemented on computers of different word-length, the deviation in round-off error could lead to different bit assignments through Equation (4.1-18). Even for the same computing equipment, the minor variation in certain arithmetic steps, for example, different logarithm evaluation for the coding and decoding operations could result in ambiguity. The ambiguity consideration is important; however, the

related difficulties can, again, be eliminate ! by careful programm-
ing. In a universal version of the coding-decoding algorithms, table
look-up should be used instead of functional evaluation, and integer
arithmetic should replace all floating point operations. The same
programming considerations also result in improved efficiency for
most general purpose computers.

The utilization of Equation (4.1-18) for sample reduction actu-
ally incorporates the novel features of both the zonal and threshold
approach to transform coding. Whenever the image power spectral
density significantly decreases for higher harmonics, Equation
(4.1-18) should lead to significant sample reduction. The coding
reduction procedure thus far outlined is highly image-dependent
(unlike zonal coding) and requires no additional bookkeeping informa-
tion (unlike threshold coding). The decoder is completely uninformed
of the degree of sample reduction; this information it can only
ascertain upon completion of the decoding process.

The natural form of the image power spectral density may not
lead to a sufficient degree of sample reduction. Appropriate applica-
tion of the filtering process in Figure 4.1-1 discussion of which
was delayed to the present, can significantly alter the bit rate.
Generally, the filter function is of the low-pass form (it can also be
image-dependent). The coder-decoder will operate on the modified
power spectrum. Thus any degree of sample reduction can be
achieved by selecting the appropriate filter. It should be observed

that, again, the application of the filter requires no bookkeeping bits to the receiver. In fact, the decoding algorithm has no information as to the type or structure of the filter if, in fact, one was utilized.

The small amount of white noise in the image significantly alters the power spectral density for the higher harmonics.[*] In fact, for the higher harmonics, the sample variances are basically specified by the noise spectral density. The coder cannot differentiate between the true image and noise. Application of the filter may lead to its more conventional role, that is, to increase the S/N of the image.

The adaptive philosophy can easily be extended to other orthogonal decompositions. The Walsh transform was utilized for this implementation. The conventional schematic representation of the sequency-ordered Walsh transform is shown in Figure 4.1-11. The Walsh transform of the 256 x 256 matrix is another 256 x 256 matrix. The "unconventional" phase concept permits the description of the transform plane by an equal number of phase and amplitude terms. The following definition was used. In Figure 4.1-11 each row is considered as 128 number pairs. These pairs are used for the amplitude and phase definition in a similar manner to the

---

[*] The effects of noise on coding are further discussed in subsection 4.2.
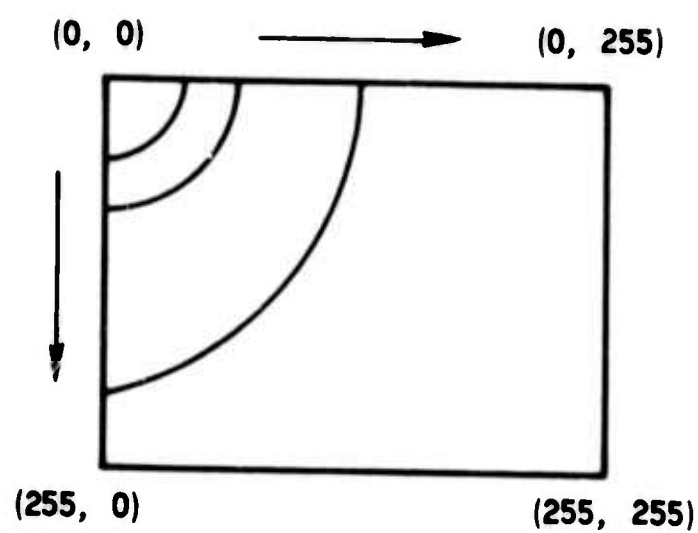
Figure 4.1-11. Conventional Walsh Domain Representation
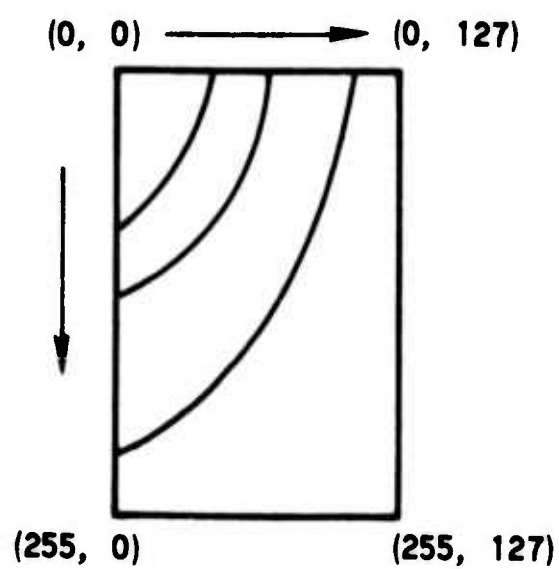
Figure 4.1-12. Walsh Amplitude Domain Representation

Fourier case. Let $(A_{i,2j-1}, a_{i,2j})$ be one such pair, then the corresponding amplitude and phase values are defined respectively as $x_{i,j} = \left(a_{i,2j-1}^2 + a_{i,2j}^2 \right)^{1/2}$ and $\tan \theta_{i,j} = a_{i,2j-1}$. The corresponding representation of the Walsh amplitudes is shown in Figure 4.1-12.

Once the amplitude plane is specified it is obvious that the various equations used for coding the Fourier plane, Equations (4.1-4) through (4.1-17), are equally appropriate. There are only two basic differences: (a) the estimator equations for the negative (left) side are not needed and (b) the sequence of operation must correspond to the symmetry of Figure 4.1-12. The coder will again proceed downward row by row. Within each row it will always proceed from the zero column to the right.

Once the coding algorithm is adjusted for the two minor differences listed above, the coding and decoding steps listed for the Fourier domain are equally valid for the Walsh domain. Similarly, the various comments relating to bit assignment, ambiguity, and implementation of computer-to-computer communications are equally valid for the Walsh coder.

4.2 Effects of Noise in the Original Image

Noise effects are considered in this section via a simplified analytic model.

For the purpose of analysis the image correlation is modeled by the simple exponential Markov expression

$$R(x, y) = e^{-\alpha |x|} \, e^{-\beta |y|} = \rho_1^{|x|} \rho_2^{|y|} \qquad (4.2-1)$$

Further simplification is obtained by the assumption of identical horizontal and vertical statistics. Therefore, let $\beta = \alpha$ and $\rho_1 = \rho_2 = \rho$, then

$$R(x, y) = e^{-\alpha(|x| + |y|)} = \rho^{|x| + |y|} \qquad (4.2-2)$$

Note also that $\alpha > 0$, $\rho < 1$, $\alpha = -\ln\rho$.

The application of the Markov model [Equation (4.2-2)] leads to interesting quantitative results. In the following, the image is assumed to be normalized such that its mean is zero and its variance is unity. It is assumed that the image is corrupted by additive white noise of power spectral density N. The average S/N in the image is, therefore, 1/N. The local S/N in the transform domain, denoted by Q, is the ratio of the image and noise power spectral densities:

$$Q(u, v) = S(u, v)/N \qquad (4.2-3)$$

Restricting the discussion to the Fourier transform, the power spectral density is given by

$$S(u, v) = \iint\limits_{-\infty}^{\infty} R(x, y) \, e^{-2\pi j(ux + vy)} dx dy \qquad (4.2-4)$$

Utilizing Equation (4.2-2) for R,

$$Q(u, v) = \frac{1}{N} \left[\frac{2\alpha}{\alpha^2 + (2\pi u)^2} \right] \left[\frac{2\alpha}{\alpha^2 + (2\pi v)^2} \right] \qquad (4.2\text{-}5)$$

Before proceeding with Equation (4.2-5) the problem must be discretized. As before, the image is assumed to be properly sampled, e.g., no aliasing, on a rectangular grid at locations

$$(x_n, y_m) = (n\Delta x, m\Delta y); \; n, m = 0, \pm 1, \pm 2, \qquad (4.2\text{-}6)$$

The appropriate frequency band limits in the transform domain are $[-1/2\Delta x, +1/2\Delta x]$ and $[-1/2\Delta y, 1/2\Delta y]$ for the horizontal and vertical directions, respectively. For computational convenience, let $\Delta x = \Delta y = 1$. Thus, both the horizontal and vertical extent of the frequency domain is $-1/2$ to $+1/2$.

The behavior of Q is considered along the diagonal in the frequency plane, e.g., $u = v = f$. Whenever the noise dominates, $Q(u, v) < 1$. Letting $Q(f, f) = 1$, one can solve for the transition region. Considerable simplification is achieved by the assumption that at $Q(f, f) = 1$, $2\pi f \gg \alpha$. The latter inequality is realistic for most images and it will be demonstrated for the specific example utilized in this section. Equation (4.2-5) can be rewritten according to the previous assumptions as

$$N = 4\alpha^2/(2\pi f)^4 \qquad (4.2\text{-}7)$$

therefore

$$f = (\alpha^2/4\pi^4 N)^{1/4} \qquad (4.2-8)$$

Since $4\pi^4 \approx 400$

$$f = \left(2.5 \times 10^{-3} \frac{\alpha^2}{N}\right)^{1/4} \qquad (4.2-9)$$

For the numerical utilization of Equation (4.2-9), α and N must be specified.   Let $N = 0.001$ and $\alpha = 0.05$ corresponding to $\rho = 0.95$ in Equation (4.2-2).   The value used for N is very conservative since it corresponds to the image S/N of 1000.   The $\rho = 0.95$ is a typical value.   The test images in this dissertation have an average sample-to-sample correlation approximately corresponding to this value.   The evaluation of Equation (4.2-9) for specified values leads to the following:

$$f = [2.5 \times 10^{-3} \times 10^3 \times (0.05)^2]^{1/4}$$

$$\approx 0.1(60)^{1/4} = 0.27 \approx \left(\frac{1}{2}\right)(0.5)$$

According to this numerical demonstration, at 1/2 of the highest vertical or horizontal harmonic the image power spectral density drops below the noise level.

Further observations are also in order.   Note should be taken that $\alpha = 0.05 \ll 0.27 \times 2\pi$, thus the simplification that led to the derivation of Equation (4.2-7) was, in fact, permitted.   One can

also note that whenever the same implication is allowed, the lines of constant S/N in the transform plane are parabolas. From Equation (4.2-5), letting $\alpha \ll 2\pi u$, $\alpha \ll 2\pi v$,

$$uv = \sqrt{\alpha^2/(4\pi r^4 NQ)} \qquad (4.2\text{-}10)$$

Equation (4.2-10) is the function of a parabola, whenever the right-hand side is a constant.

The maximum value of Q is at the $u = v = 0$ location in the frequency plane. It is

$$Q(0,0) = 4/\alpha^2 N = 1.6 \times 10^6 \qquad (4.2\text{-}11)$$

for the previously-specified values of α and N. The demonstrated example indicates that the presence of even a small amount of white noise will have a very significant effect. In this example, the majority of transform domain samples are below the noise level despite the fact that the noise level is approximately six orders of magnitude below the peak of the power spectral density.

The previous analysis can be easily extended to the case where the image correlation model is isotropic. For this case

$$R(x,y) = R\left(\sqrt{x^2 + y^2}\right) = e^{-\alpha\sqrt{(x^2 + y^2)}} \qquad (4.2\text{-}12)$$

Letting $(x^2 + y^2)^{1/2} = r$, and $(u^2 + v^2)^{1/2} = f$, Equation (4.2-4) is replaced by the Hankel transform:

$$S(f) = \int_0^\infty 2\pi r \, R(r) \, J_0(2\pi r) \, dr \qquad (4.2\text{-}13)$$

and thus

$$Q(f) = \frac{1}{N} (2\pi\alpha) \left[\alpha^2 + (2\pi f)^2\right]^{-3/2} \qquad (4.2\text{-}14)$$

as in the derivation of Equation (4.2-7) the inequality $2\pi f \gg \alpha$ can be used, therefore whenever $Q(f) = 1$,

$$N = \frac{\alpha}{(2\pi)^2} f^{-3} \qquad (4.2\text{-}15)$$

Furthermore,

$$f = \left[\frac{\alpha}{(2\pi)^2 N}\right]^{1/3} \qquad (4.2\text{-}16)$$

The utilization of the previously-specified parameters ($\alpha = 0.05$ and $N = 0.001$) indicates that $f > 1/2$, therefore the transform domain S/N in this instance will remain above unity in its region of definition.

Although both of the previous models can be expected to deviate from the actual image power spectral density, the qualitative results are useful in that they demonstrate the importance of image noise.

4.3   Pictorial Examples

The coding procedures previously outlined in this chapter were programmed for computer implementation. The results, using the

monochrome test images of Appendix A, are shown in this sub-section.

The logarithmic amplitude and linear phase displays are shown in Figure 4.3-1 for the Fourier and Walsh transforms. The same figure includes a demonstration of entropy associated with these phase images. To make the visual comparison between the two transforms more meaningful, the conventional Walsh presentation (Figure 4.1-11) is remapped and is shown according to the schematics of the conventional Fourier display. All pictorial Walsh domain presentations in this dissertation are done in this manner. The entropy images are obtained in two steps. First, the phase range $[-\pi, \pi]$ is uniformly quantized by a 64-level (6 bit) quantizer. Next, the two-dimensional probability density function (e.g., histogram) corresponding to the simultaneous occurrence of phase values corresponding to adjacent row neighbors is calculated. The actual entropy map is obtained by taking the base 2 logarithm of the two-dimensional histogram. The actual entropy value corresponding to this map is obtained by summing all 4096 elements and it is 11.99 bits (the maximum possible is 12 bits) for both transforms. The obvious conclusion is that the various phase values are, in fact, uncorrelated. The higher intensity level along the phase image diagonals does indicate a small amount of residual correlation.

The processed images are shown in Figures 4.3-2 and 4.3-3 corresponding to the Fourier and Walsh transforms, respectively.

(a) Amplitude (Fourier)
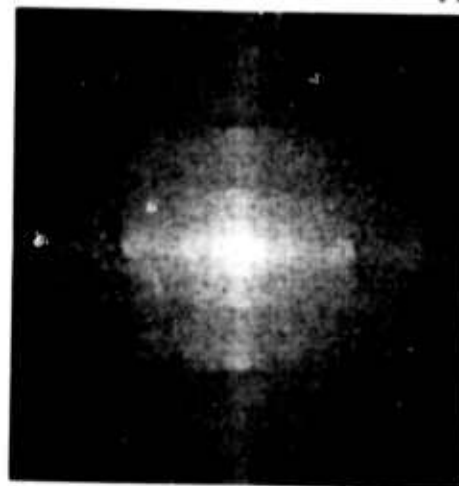
(b) Amplitude (Walsh)
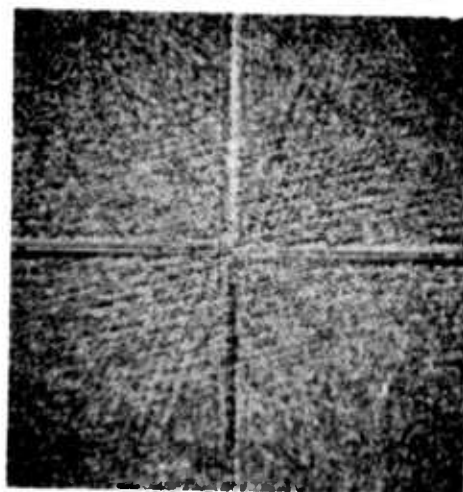
(c) Phase (Fourier)
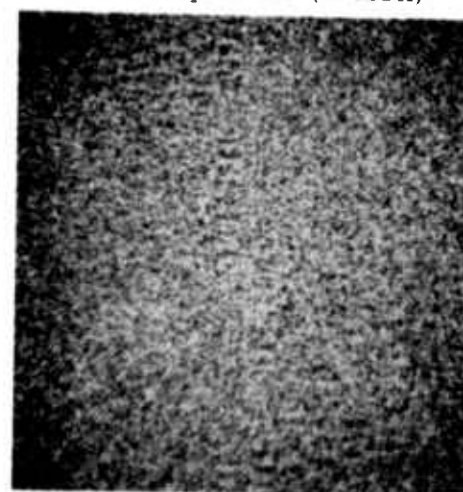
(d) Phase (Walsh)

(e) Entropy (Fourier Phase)

(f) Entropy (Walsh Phase)

Figure 4.3-1. Transform Domain Display of GIRL Image

(a) M.S.E. 2.4%; bit rate 0.39   (b) M.S.E. 1.6%; bit rate 0.68

(c) M.S.E. 1.26%; bit rate 0.31   (d) M.S.E. 0.78%; bit rate 0.66

Figure 4.3-2. Coding-Decoding Examples (Fourier Transform).
The Mean Squared Error (M.S.E.) is Normalized
Relative to Energy in Original Image

(a) M.S.E. 3.6%; bit rate 0.51

(b) M.S.E. 2.6%; bit rate 0.75

(c) M.S.E. 1.48%; bit rate 0.50

(d) M.S.E. 1.07%; bit rate 0.73

Figure 4.3-3. Coding-Decoding Examples (Walsh Transform).
The Mean Squared Error (M.S.E.) is Normalized
Relative to Energy in Original Image

In the coding-decoding process, the image is always converted into a sequence of binary digits corresponding to the actual data rate. The decoder uses the same sequence as its input.

In Figure 4.3-4, the decoded transform planes are shown. For this case, the sample reduction is obtained by rotationally symmetric low-pass filtering. It should again be stated that the decoder is uninformed about the type, or even the existence, of this low-pass filter.

Typical examples of the "dynamically" determined bit planes are shown in Figures 4.3-5 and 4.3-6.

Typical performance curves are shown in Figure 4.3-7. The various curves were generated according to the following procedure. Let $\widetilde{I}(\rho, \theta)$ and $\widehat{\widetilde{I}}(\rho, \theta)$ represent the original and decoded image transforms in polar coordinates; also the normalization relative to integrated variance is assumed:

$$\int_0^{2\pi} \int_\epsilon^\infty \widetilde{I}(\rho, \theta) \, \rho \, dp \, d\theta = 1 \qquad (4.3-1)$$

The lower limit ϵ indicates that the d.c. term is excluded in the integration. The letter designations a through e correspond to the following five functions designated as Z_a through Z_e, respectively:

$$Z_a(\rho) = \int_0^{2\pi} \int_\epsilon^\rho \left| \widetilde{I}(s, \theta) \right|^2 s \, ds \, d\theta \qquad (4.3-2)$$
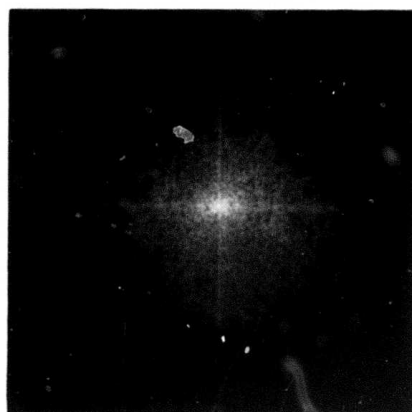
(a) Fourier Amplitude
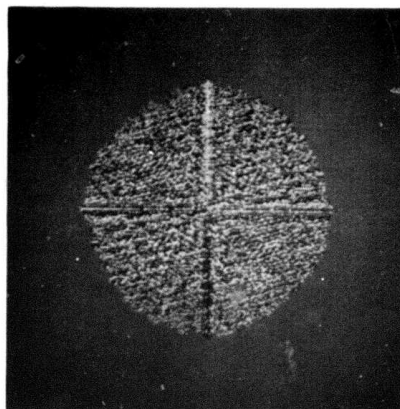
(b) Fourier Phase
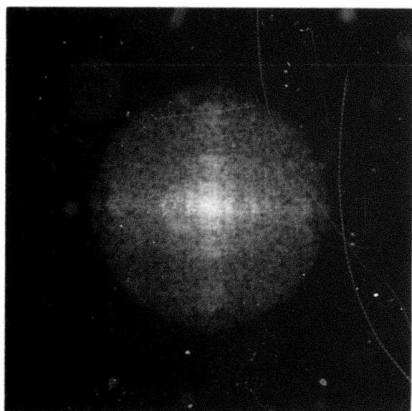
(c) Walsh Amplitude
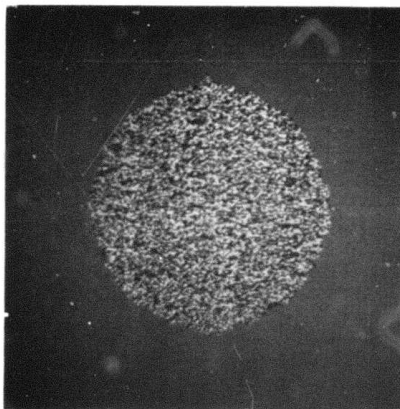
(d) Walsh Phase

Figure 4.3-4.   Examples of Decoded Transform Planes for GIRL Image

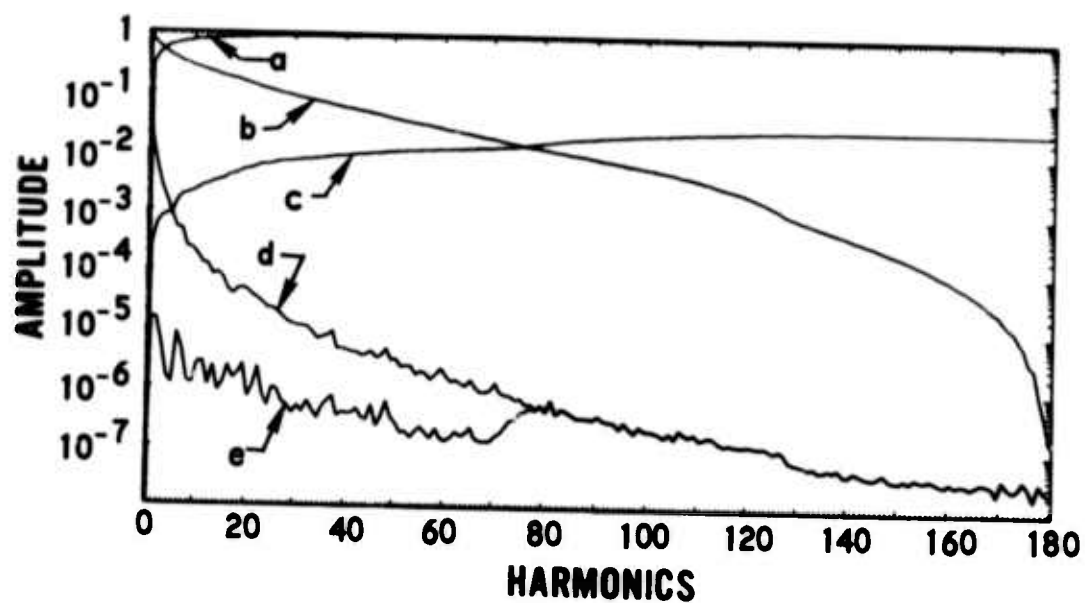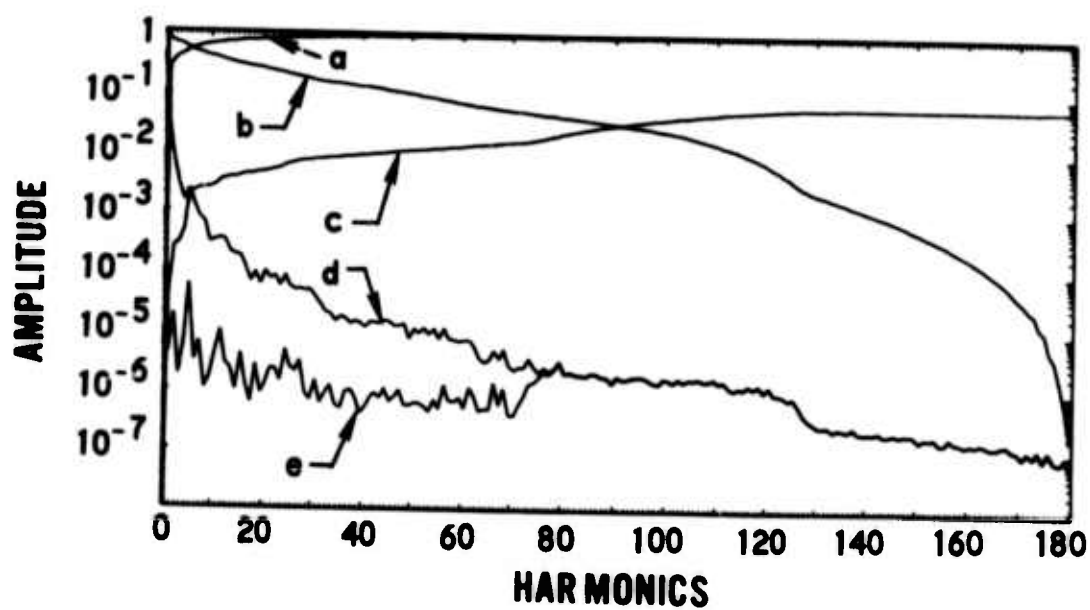Figure 4.3-5. Example of Fourier Amplitude Bit Allocation
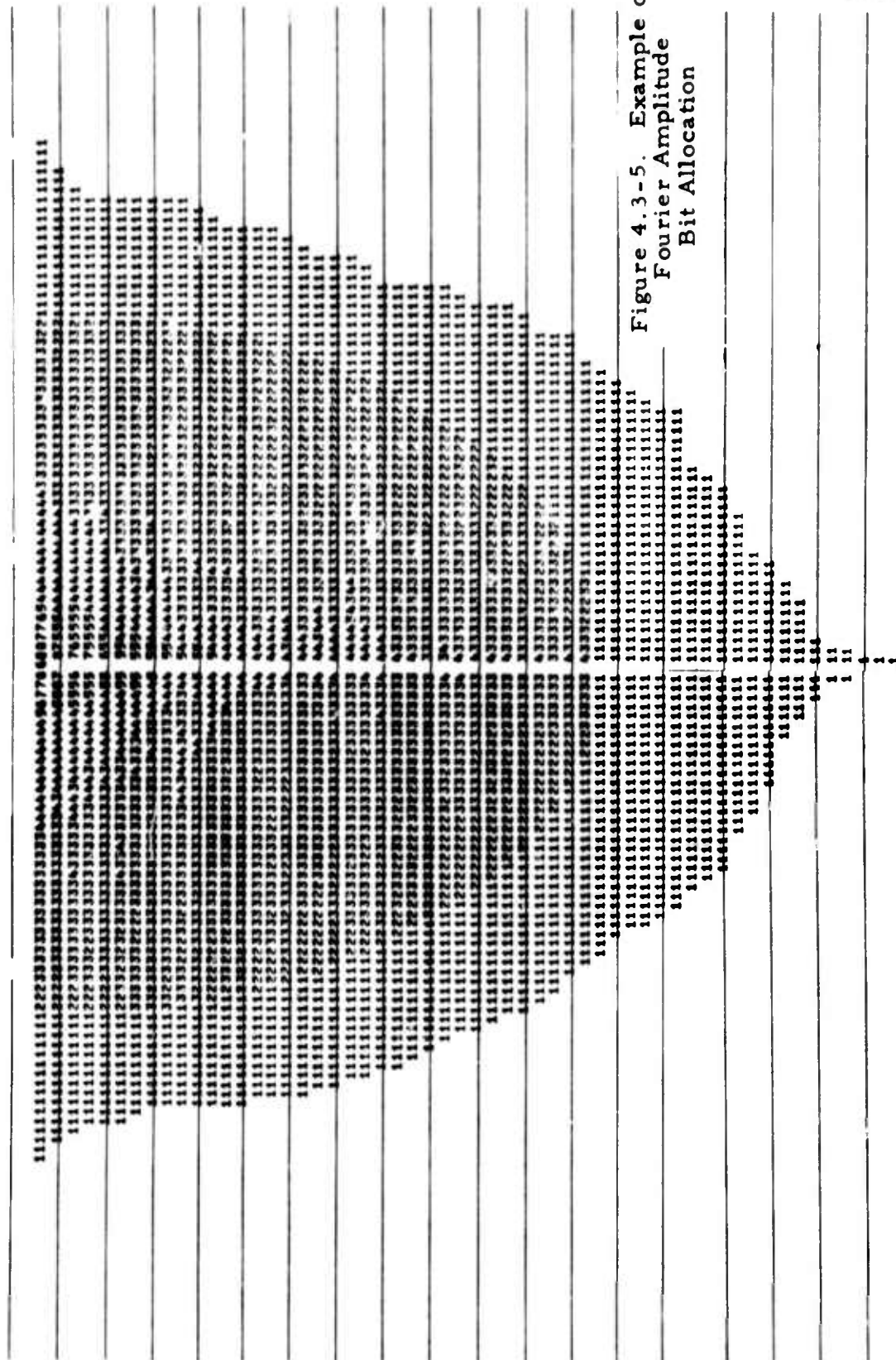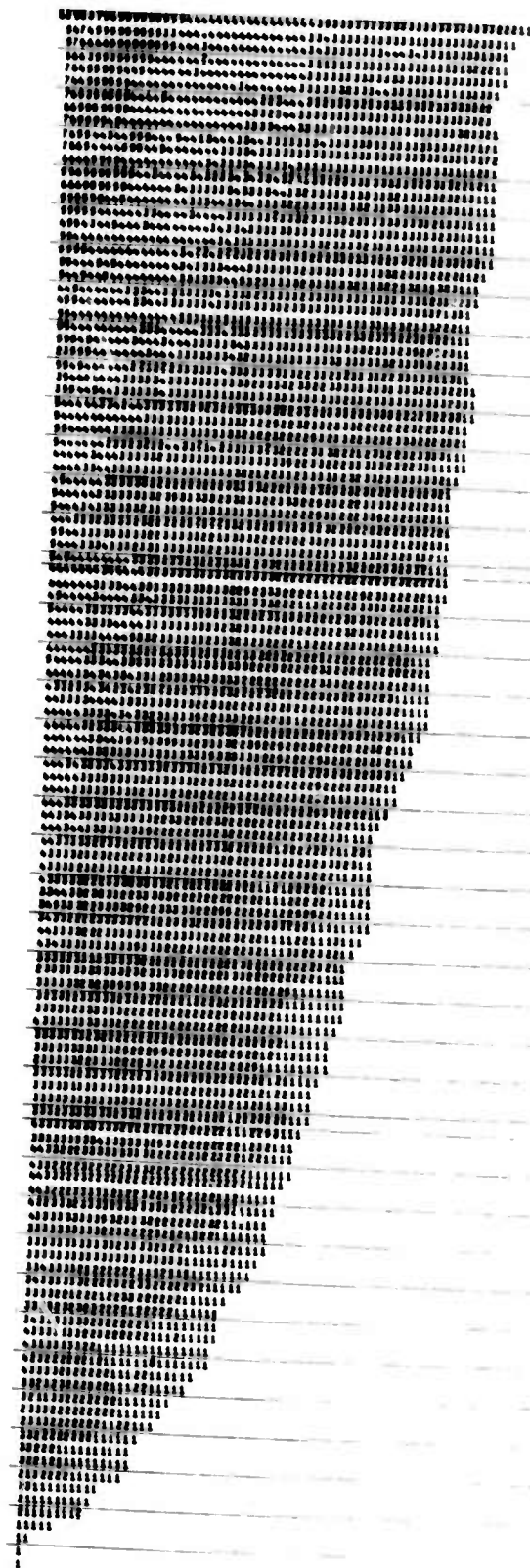
Figure 4.3-6. Example of Walsh Amplitude Bit Allocation

(a) Fourier Domain

(b) Walsh Domain

Figure 4.3-7. Typical Performance Curves for Coding-
Decoding Examples (see text for various
letter designations)

$$Z_b(\rho) = 1 - Z_a(\rho) \tag{4.3-3}$$

$$Z_c(\rho) = \int_0^{2\pi} \int_\epsilon^\rho \left| \widetilde{I}(s,\theta) - \widehat{\widetilde{I}}(s,\theta) \right|^2 s\,ds\,d\theta \tag{4.3-4}$$

$$Z_d(\rho) = \frac{1}{2\pi\rho} \frac{d}{dp} Z_a(\rho) \tag{4.3-5}$$

$$Z_e(\rho) = \frac{1}{2\pi\rho} \frac{d}{d\rho} Z_c(\rho) \tag{4.3-6}$$

The curves in Figure 4.3-7 were generated from the discretized versions of Equations (4.3-2) through (4.3-6). These functions convey considerable information about the coding process (although in a forced rotational symmetry). Z_a is the integrated transform variance. Z_b corresponds to the truncation error. The integrated overall coding error is Z_c. The image power spectral density is Z_d. The local (in the transform plane) coding error is given by Z_e. The curves Z_d and Z_e merge at the location of the low-pass filter boundary.

In Figure 4.3-8, the sample reduction is obtained by discarding transform samples whenever the amplitude is below a certain value. The subsequent coding is the same as in previous examples.

The influence of apodizing, a ten-element tapered window in this case, is demonstrated in Figure 4.3-9.

(a) M.S.E. 1.83%; bit rate 0.6

(b) M.S.E. 1.18%; bit rate 0.42
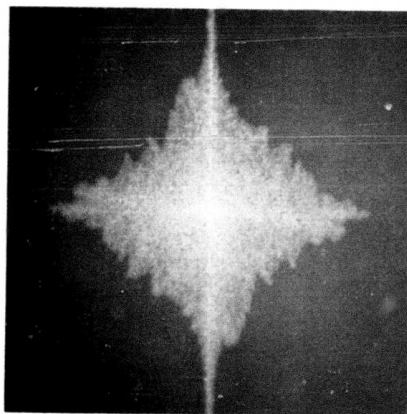
(c) Decoded Transform Plane
for COUPLE IMAGE
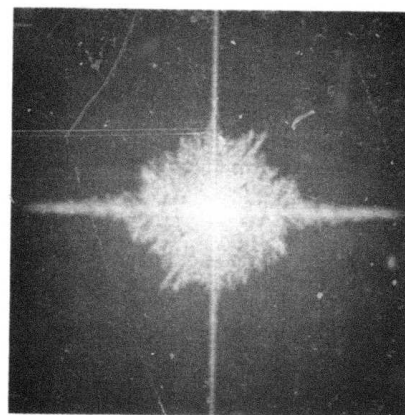
(d) Decoded Transform Plane
for GIRL Image

Figure 4.3-8. "Threshold" Coding Experiment (Fourier Transform). The Mean Squared Error (M.S.E.) is Normalized Relative to Energy in Original Image
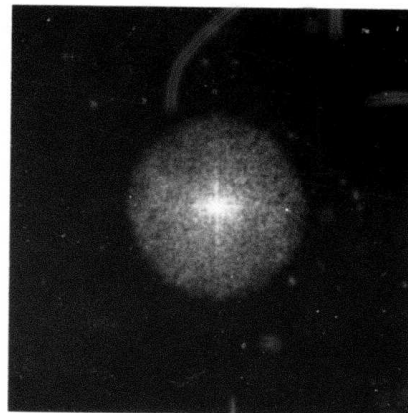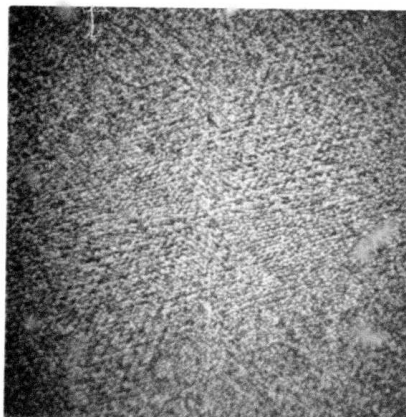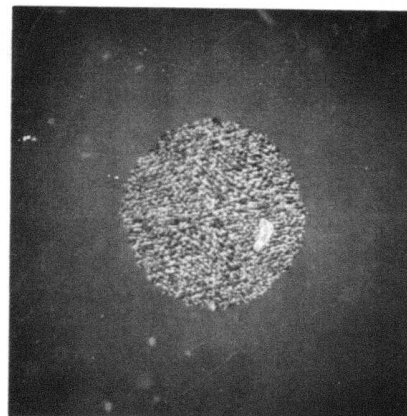
(a) Decoded Image (0.4 bit)

(b) Original Amplitude

(c) Decoded Amplitude

(d) Original Phase

(e) Decoded Phase

Figure 4.3-9.   Coding Experiment with Apodizing (Fourier)

4.4 Channel Coding Considerations

There are two basic philosophies concerning channel noise sensitivity evaluation of the picture coder. In one case, channel errors are permitted to occur in the source coded data and, consequently, the reconstructed image will be affected by the channel noise. A "well-behaved" coding process will be insensitive to channel errors. The second approach implies the requirement for channel coding and in effect assumes that by proper channel coding error-free transmission is possible. The author is a strong believer in the latter philosophy.

Lack of sensitivity to channel errors is a desirable image coding feature. It is easy to demonstrate, however, that, in general, efficient data compression and insensitivity to channel errors are contradictory concepts. The fundamental theoretical basis for any data compression procedure is redundancy in the data source. The fact that the source output is correlated permits representation of the source in "compressed" form. An efficient data compression procedure removes the existing source correlation and produces an output which, by design, will be uncorrelated. In the binary representation of the compressed data, each bit will acquire an added importance and its reversal is more apt to degrade the quality of the reconstructed data than a similar occurrence of error in the original (uncompressed) data.

It is not surprising that most efficient image coding algorithms, particularly contour coding techniques, are very sensitive to channel errors. For the latter method, a single bit error is likely to prevent the entire image reconstruction.

Source coding removes the source redundancy, conversely, redundancy is reintroduced by channel coding. For the channel coding procedure, it is important that the input to it be of a particular statistical structure. Since the channel is unlikely to have been designed to accommodate any particular source redundancy, it is anticipated that channel usage is optimum when its input is statistically uncorrelated.

According to the above statement, source redundancy, i.e., finite memory, is undesirable for subsequent channel coding which assumes a memoryless source. The PCM form of the image is highly correlated. The high degree of correlation can be demonstrated in the binary equivalent of the image.

A quantitatively meaningful demonstration of correlation is the correlation function calculated from the binary equivalent of image segments. Each value of the correlation function R_j is determined from a data segment of N values:

$$R_j = \frac{1}{N-j} \sum_{i=1}^{N-j} (x_i - \bar{x})(x_{i+j} - \bar{x}) \qquad (4.4-1)$$

and

$$\bar{x} = \sum_{i=1}^{N} x_i \qquad\qquad (4.4\text{-}2)$$

The data (x_i) are the actual binary representation of the image, that is, each x_i is either 1 or 0. Calculation of the correlation function for $N = 10,000$ from the PCM "Girl" image is shown in Figure 4.4-1a. The structural form of R_i in Figure 4.4-1a is consistent with the eight-bit representation, each relative maximum occurs at multiples of eight. R_0 is the variance of the binary stream.

The similar calculation of the correlation function for the source coder output of Chapter 4 is shown in Figure 4.4-1b. The result indicates that the source coding algorithm output is equivalent to a memoryless source.

Lack of correlation and the significant bit reduction in the output of the image coding procedure indicate that the compressed data are expected to be sensitive to channel errors. This sensitivity is evident also upon careful examination of the coding process of Section 4.1. Both the number and the location of quantum levels are "dynamically" determined. Channel error can affect both of these quantization parameters. Erroneous determination of the number of quantum levels is a catastrophic error. The synchronization in the adaptive coding procedure will be lost and all subsequent values will be drastically altered.
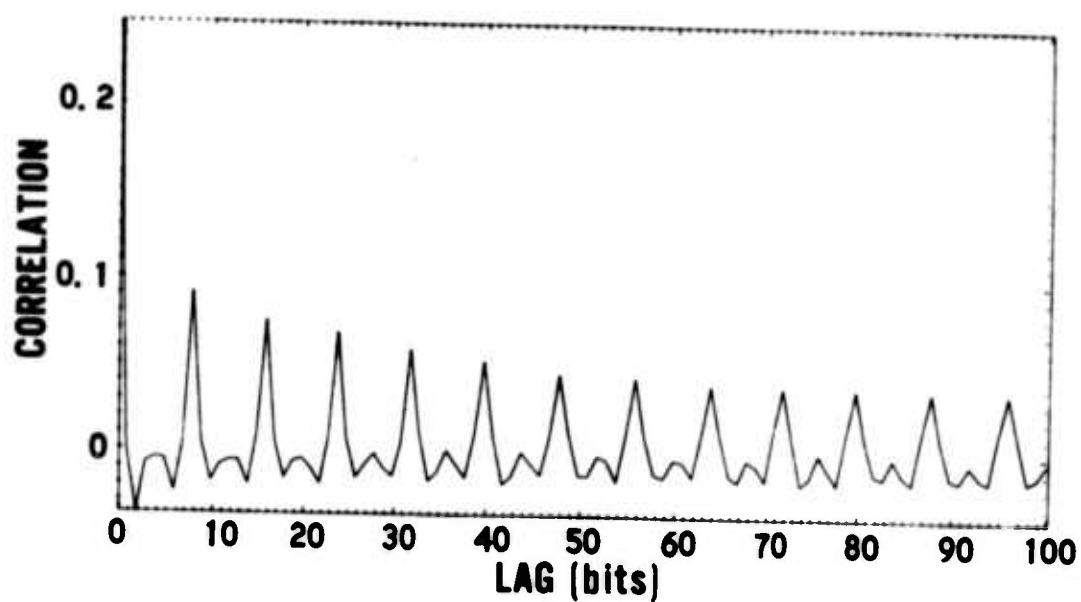
(a) Original PCM
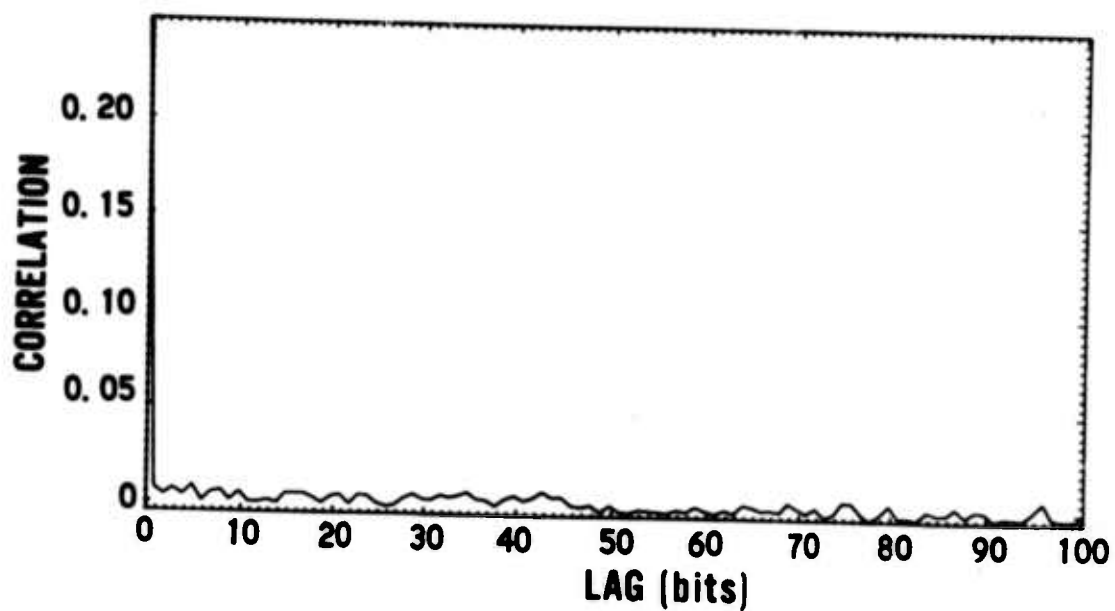
(b) Adaptive Transform Code

Figure 4.4-1. Typical Bit Stream Correlation Properties
of GIRL Image

It can be seen that the adaptive phase coder output should not be transmitted over noisy channels without channel coding. Although the degree of image compression depends on the image correlation via the generalized power spectral density, it can be controlled to virtually any reasonable value by an appropriate linear filter. Thus, the additional bandwidth requirement for the channel coder-introduced redundancy can easily be offset by the image coder at the expense of additional low-pass filtering.

By relaxing the requirement for complete adaptivity the coding algorithm can be made less sensitive to channel errors. The prior specification for bit assignment according to some conventional models retains the flexibility for determination of quantum levels, however, catastrophic errors resulting from loss of synchronization can no longer occur.

Using the polynomial surface fit for the optimum quantization parameters (subsection 3.7) would also avoid catastrophic errors, providing that the transmission of the appropriate coefficients is without errors. Neither technique was, however, employed. In either case, the additional complexity could be avoided by proper channel coding, which probably would require less total effort.

5. EXPERIMENTAL RESULTS, II (COLOR)

Transform coding techniques for monochrome images have been successfully utilized by various researchers. For other, more complex, types of images, redundancy exists in parameters other than the two spatial variables. This chapter considers the extension of the algorithm of Chapter 4 to color images, while in Chapter 6 the coding algorithm implementation is for a sequence of time-varying images.

Further extensions could include the simultaneous consideration of color and time, however, it was not done here. It should be emphasized that for even the three-dimensional data, e.g., color, or time-dependent images, the experimental difficulties become quite significant. The generation and calibration of properly registered frames is a major effort by itself. Similarly, the display and recording of a color image requires a great deal of additional hardware and care as compared to monochrome images. Furthermore, the third dimension significantly increases the data handling. The experimental difficulties listed above have kept research on color and frame-to-frame coding at a fraction of the effort extended to the monochrome case.

5.1 Color Image Representation

A passive, opaque (non-emitting) object becomes "visible" by reflecting radiation which is incident on it. The reflection process is selective, thus, the relative amount of reflected energy is

dependent on the local characteristics of the object and it also depends on the spectral distribution of the incident radiation. The physical reflection process, in effect, specifies the image.

The spectral dependence is "integrated out" for monochrome images. The exact characterization of actual visual scenes requires the specification of the spectral component which is accomplished by the third image variable, the wavelength (λ). Symbolically, the color image is an analog function of three variables, $I(x, y, \lambda)$. Prior to the coding procedure, the image must be sampled along the spectral axis in addition to the discretization of the two spatial coordinates.

The sampling procedure applied to the wavelength very strongly depends on the ultimate purpose for which the image was recorded. Formally, the spectral sampling can be written

$$I(x, y, j) = \int_0^\infty r_j(\lambda)\, I(x, y, \lambda)\, d\lambda, \quad j = 1, \ldots N \quad (5.1\text{-}1)$$

The "spectral aperture," $r_j(\lambda)$ determines the weighting of the spectral components for the determination of the j-th sample. The number of samples, N, depends on the application. Equation (5.1-1) can represent the monochrome image of Chapter 4 by specifying $N = 1$ and r_1 to be a constant over the visible portion of the spectrum.

The value of N may be in excess of 20 for what is generally referred to as multispectral data. The functional form of r_j, in this case, is usually an approximate delta function centered at a specific wavelength value, λ_j. Not all λ_j's are necessarily in the visible

spectral region. Multispectral data are generally utilized in computer classification algorithms rather than for the actual reproduction of the visual scene for human viewing.

The spectral sampling is greatly simplified for the case when the purpose of image recording is for subsequent display for human visual viewing. The human eye does not respond individually to the infinite number of spectral elements present in a visual scene. It is rather a triplet of photoelectric detectors whose individual responses cover the low (red), medium (green) and high (blue) spectral regions. The human visual process determines the color on the basis of the simultaneous "readings" of these detectors (Cornsweet, 1970).

In effect, the human eye perceives the complex visual scenes corresponding to its three detectors. Within this somewhat over-simplified model the eye performs the mapping of the continuous wavelength axis into a set of three values. The mapping is of the form of Equation (5.1-1) and it is given by the following three equations.

$$R(x,y) = \int_0^\infty r_R(\lambda)\, I(x,y,\lambda)\, d\lambda \qquad (5.1-2)$$

$$G(x,y) = \int_0^\infty r_G(\lambda)\, I(x,y,\lambda)\, d\lambda \qquad (5.1-3)$$

$$B(x,y) = \int_0^\infty r_B(\lambda)\, I(x,y,\lambda)\, d\lambda \qquad (5.1-4)$$

Letter indices R, G, and B designate the spectral region in which the appropriate "eye detector" reaches its maximum response (red, green, and blue, respectively). Equations (5.1-2) through (5.1-4) imply the nonunique spectral sensitivity of the human eye. According to these equations, the change of $I(x, y, \lambda)$ will not be perceived as long as the left sides of these equations are not altered.

From the standpoint of this bandwidth reduction algorithm, the coding-decoding process will simultaneously consider the three image signals $R(x, y)$, $G(x, y)$, and $B(x, y)$. The redundancy reduction is achieved by considering the correlation among the three color planes in addition to the spatial correlation within each plane.

5.2 Description of the Algorithm

The coding scheme for the multidimensional data which is presented in this dissertation can be put into the simple form, shown in Figure 5.2-1, in a manner similar to the monochrome case of Figure 4.1-1.

The three-dimensional transform of the R, G, B planes yields three transform planes $\widetilde{I}_1$, $\widetilde{I}_2$, $\widetilde{I}_3$. By assumption, the samples are uncorrelated within each transform plane as well as among the various transform planes.

The actual implementation utilized the Fourier transform. The three-dimensional transform is performed in two stages. The conventional two-dimensional Fourier transform is applied to the R,
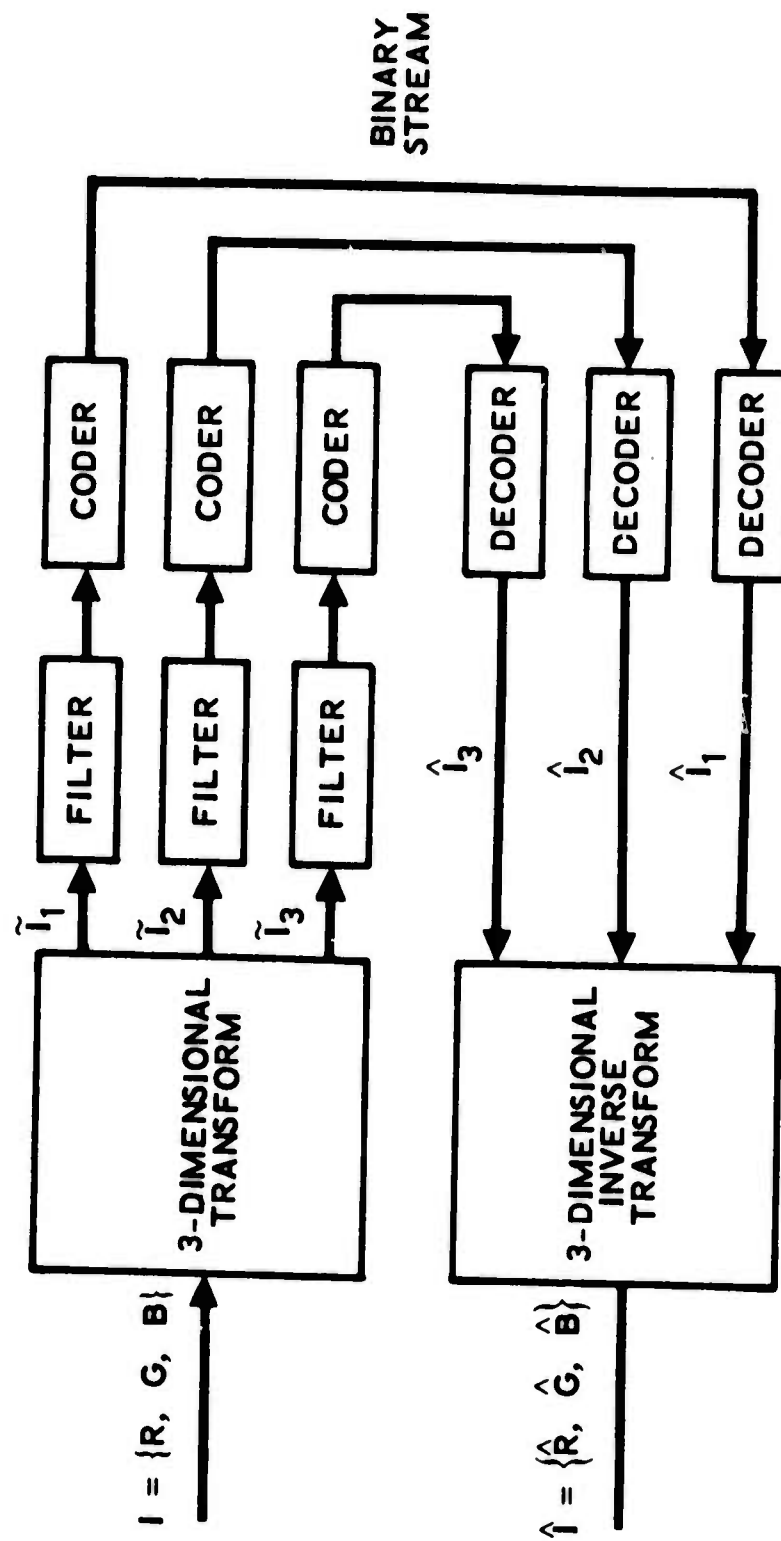
Figure 5.2-1. Schematics of Coding-Decoding Process for Color Image

G, B planes individually. Each triplet of complex values corresponding to identical locations in the three transformed planes is subjected to the one-dimensional, three-point complex Fourier transform. The 3 x 3 complex unitary transform matrix is shown in Figure 5.2-2.

$$\frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0.5\,(-1 + j\sqrt{3}) & -0.5\,(1 + j\sqrt{3}) \\ 1 & -0.5\,(1 + j\sqrt{3}) & 0.5\,(-1 + j\sqrt{3}) \end{bmatrix}$$

Figure 5.2-2. Three-Element Fourier Transform Matrix

The structure of the 3 x 3 complex matrix indicates that the first of the three final transform planes $(\tilde{I}_1)$ is simply the average of the two-dimensional transform of the R, G, B planes, respectively. In effect, the first plane, $\tilde{I}_1$, contains the brightness, or luminance, information. The $\tilde{I}_2$ and $\tilde{I}_3$ planes designate the fluctuation around the average of the three planes and thus represent the chrominance information. It is not necessary, however, to make reference to luminance and chrominance designations in order to implement the coding procedure.

The coding-decoding procedure of Chapter 4 is applied to the three transform planes $(\tilde{I}_1, \tilde{I}_2, \tilde{I}_3)$ individually. The only interdependence among the three separate coding processes is that the

scale factor that relates the number of quantization levels to the sample variance is determined for $\widetilde{I}_1$ and the same value is utilized for $\widetilde{I}_2$ and $\widetilde{I}_3$ as well.

As in the procedure utilized for the monochrome case in Chapter 4, the three filters can be used by the transmitter to modify the coding process. The receiver will adapt to the filtered planes without any control information. The close similarity of the R, G, B and their two-dimensional transforms implies that the largest image energy component will be concentrated on the first transform plane.

The low-pass filtering effects of the $\widetilde{I}_1$, $\widetilde{I}_2$, and $\widetilde{I}_3$ planes are similar to the procedure previously used for the Y, I, Q system (Pratt, 1971). The "luminance" plane $\widetilde{I}_1$, in effect represents the spatial resolution, which will be degraded by a high degree of low-pass filtering. The "chrominance" planes $\widetilde{I}_2$ and $\widetilde{I}_3$ can be subjected to rather strong low-pass filtering without serious image degradation. The replacement of every value by zero in the $\widetilde{I}_2$ and $\widetilde{I}_3$ planes reduces the color image to a monochrome equivalent. This mono-chrome image is simply the average of the R, G, and B signals.

The value of the adaptive nature of the color coding process as indicated in Figure 5.2-1 cannot be overemphasized. The appropriate filters can be specified for a specific color system. For the general case, the R, G, B signals may be referenced to a wide variety of primaries. The degree of low-pass filtering which may

be tolerated will depend on the original color digitizing equipment
and the subsequent calibration procedure, if any. The low-pass
filter band limits are implemented by the transmitter and the spec-
ification is very likely the result of human visual inspection. When-
ever the cost in effort associated with color image transmission is
high, optimization of the three filters is likely to require consider-
able effort.

The significant property of the adaptive procedure is that once
the transmitter decides on an optimized set of filters (i.e., the
tolerance of the transform planes to low-pass filtering has been
determined) none of this information is required by the receiver.
All information bits relate to the quantized transform domain, and
no bookkeeping information is required.

The same comments made in Chapter 4 regarding advantages
and disadvantages apply for the adaptive color coder as well. The
adaptive procedure includes the benefits of both zonal coding (non-
uniform bit assignment and quantum levels) and threshold coding
(adaptivity in deciding which regions can be discarded). The funda-
mental disadvantage of the adaptive coder is the variable buffer
requirement for the receiver. The bit rate, or equivalently, the
degree of compression, is determined by the transmitter and only
after decoding will this information be available to the receiver.

The three-primary color system utilized the three-dimensional
Fourier decomposition only. The one-dimensional, 3-point Walsh

transform is not defined. An alternate composite system could include the two-dimensional Walsh transform of the R, G, B planes followed by the one-dimensional Fourier transform. The alternate system was not actually implemented. If the number of the color planes N is of the form $N = 2^n$, where n is an integer, the Walsh decomposition is possible for the third dimension. The three-dimensional Walsh decomposition of Chapter 6 could be directly utilized for multispectral data of four input planes.

5.3 Pictorial Examples

As with the monochrome coder of the previous chapter, the color coding algorithm has been programmed. The three-dimensional Fourier transform was utilized.

Figure 5.3-1 shows the three transform planes corresponding to the three-dimensional transform. Results of the coding-decoding experiments are shown in Figure 5.3-2. Tristimulus color planes for one of the decoded images are shown in Figure 5.3-3. Decoded transform planes for a typical case are shown in Figure 5.3-4. This figure also indicates the varying amount of low-pass filtering in the different transform planes. The transform statistics can be significantly altered by apodizing. The influence of a ten-element image window is demonstrated in Figure 5.3-5.

It should be noted that the d.c. term for the three-dimensional transform is located in the first transform plane. For the other transform planes, the relative maximum amplitude location is not predictable.
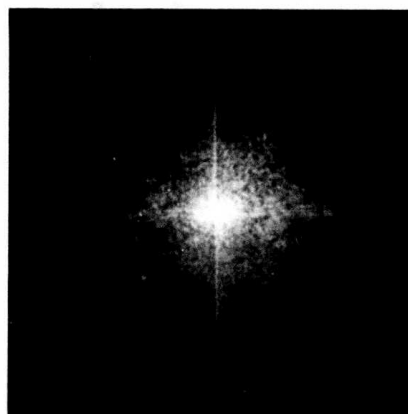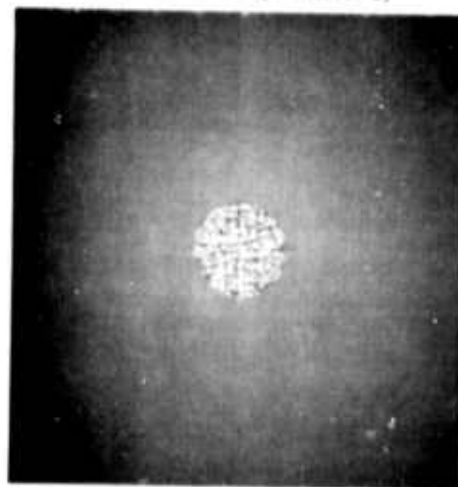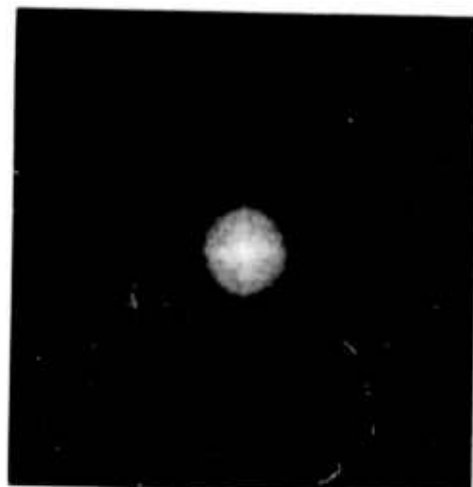
(a) Amplitude (Plane I)

(b) Phase (Plane I)
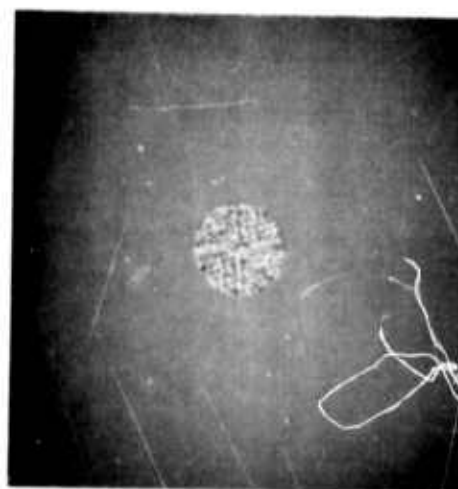
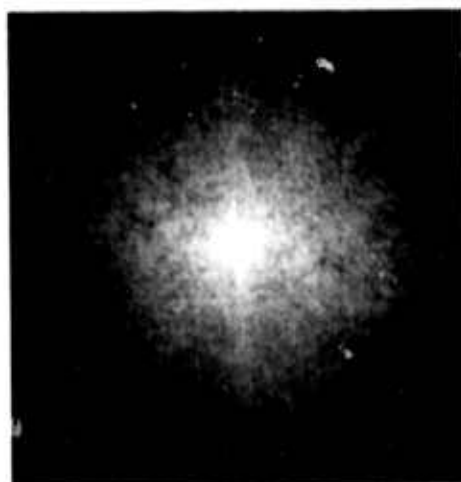(c) Amplitude (Plane II)
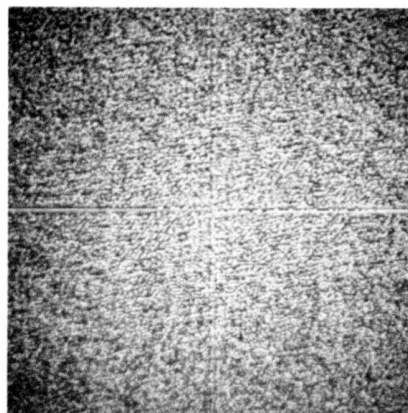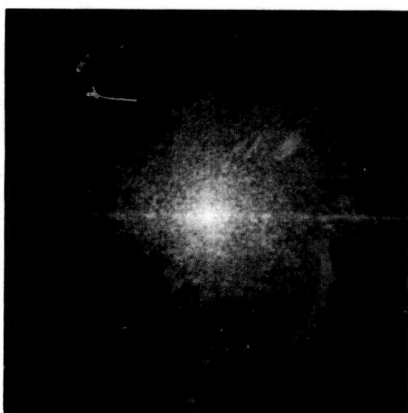
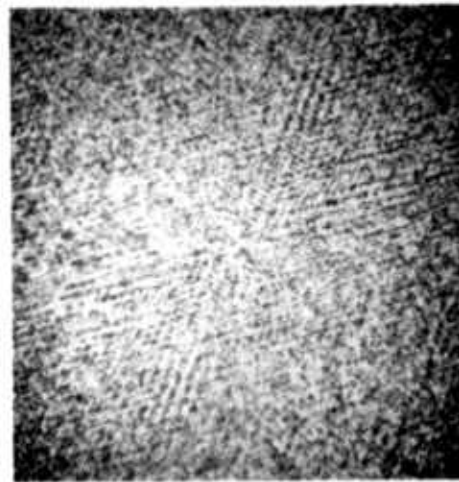(d) Phase (Plane II)

(e) Amplitude (Plane III)
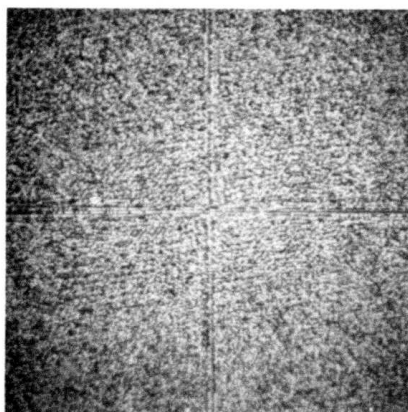
(f) Phase (Plane III)

Figure 5.3-1.   Three-Dimensional Fourier Transform
Display of Color GIRL Image

(a) Apodized; Bit Rate 0.65

(b) M.S.E. 1.9%; Bit Rate 0.55

(c) M.S.E. 1.2%; Bit Rate 1.19

(d) M.S.E. 4.5%; Bit Rate 0.54

(d) M.S.E. 2.4%; Bit Rate 1.2

Figure 5.3-2. Coding-Decoding Examples. The Mean Square Error (M.S.E.) is Normalized Relative to Energy on Original Color Image

(a) $\hat{R}$

(b) $\hat{G}$

(c) $\hat{B}$

Figure 5.3-3. Tristimulus Color Planes of Decoded
GIRL Image (0.62 bit)

(a) Amplitude (Plane I)

(b) Phase (Plane I)

(c) Amplitude (Plane II)

(d) Phase (Plane II)

(e) Amplitude (Plane III)

(f) Phase (Plane III)
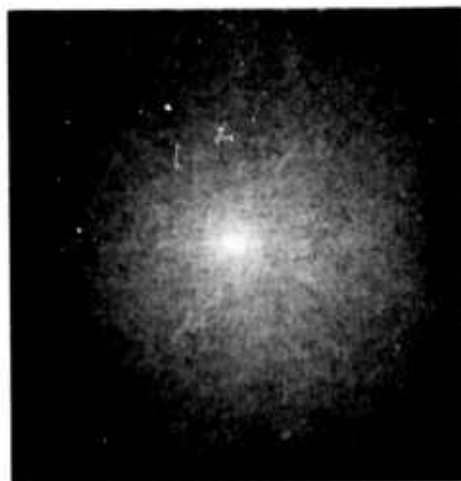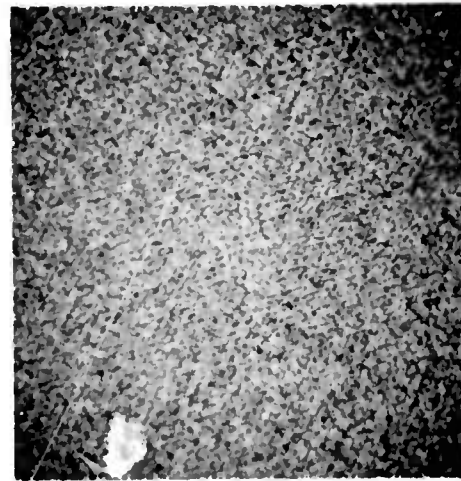
Figure 5.3-4. Decoded Transform Planes for Color GIRL
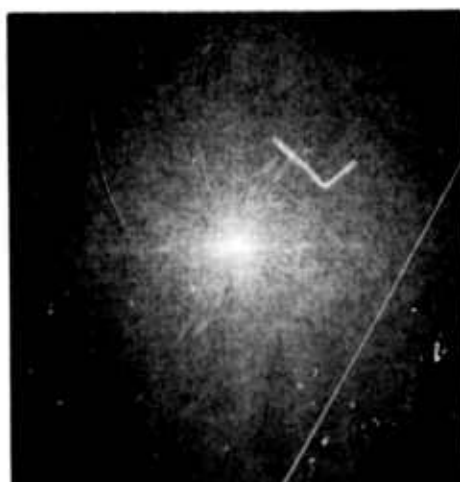
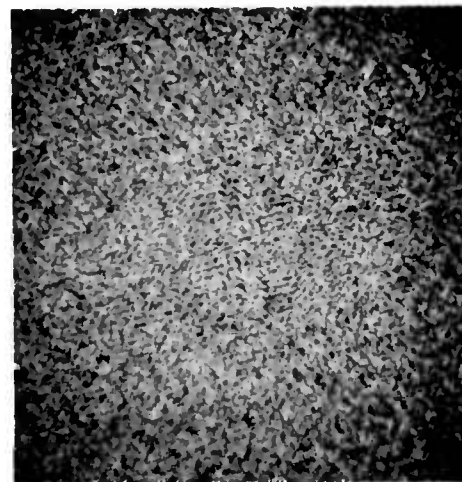(a) Amplitude (Plane I)

(b) Phase (Plane I)

(c) Amplitude (Plane II)

(d) Phase (Plane II)

(e) Amplitude (Plane III)

(f) Phase (Plane III)

Figure 5.3-5. Transform Domain Display Associated with Apodized Color Girl Image

6. EXPERIMENTAL RESULTS III (INTERFRAME CODING)

It is obvious that significant redundancy exists among members of image sequences representing the temporal variation of visual scenes. An efficient interframe coder removes the redundancy in the sequence of similar images as well as within each image (Haskell, Mounts, and Candy 1972).

In a manner similar to the color coding approach of the previous chapter, the image sequence can be considered as three-dimensional data consisting of two spatial coordinates and one time coordinate. An additional similarity between interframe and color coding is that they both exploit the limitation of the human visual system. For most practical applications, three primary color components are sufficient to represent most apparent colors within the spectral range of the human visual response. The limited temporal resolution of the human eye permits the sampling of the temporal variable at approximately 60 Hertz.

It is important to note that the sampling procedure thus specified by the inadequacy of the human visual process does not necessarily correspond to the classical sampling requirement. Consequently, emphasis by the coder is on preservation of the appearance of the image rather than on the actual image itself.

There are many applications for interframe coding, the most obvious being television. Although various spatial domain techniques have been successfully utilized for redundancy reduction in video signals, transform techniques have not been previously considered.

143

This chapter considers the theoretical implications for transform domain coding. The adaptive phase coding technique of Chapter 4 is extended to the interframe case. Detailed discussion of the algorithm as well as examples of the coding procedures are also given.

6.1 Analysis of the Interframe Case

A sequence of images can formally be written as $I(x, y, t)$. The spatial variation is indicated by x and y and the temporal variation by t. The three-dimensional function I represents the continuous (non-discrete) variation of a visual scene. The physical nature of the imaging process requires that $I(x, y, t)$ be non-negative.

If I is band-limited with respect to all three of its variables, then the Lukosz bound applies, at least formally. In fact, the discussion in subsection 2.6 indicated the tightening of this bound for increasing numbers of dimensions. If the Lukosz bound is to be valid for the sampled version of I, the sampling rate must be at least twice the band limit for each dimension. The various imaging devices band-limit the spatial frequency spectrum of images; however, no similar band-limiting occurs for the temporal variation. Further, sampling along the time axis is performed to match the limitations of human visual process and bears no relation to the structural form of the actual image. Consequently, the Lukosz bound does not apply for the temporal portion of interframe imagery.

The utility of statistical coding should also be discussed as applicable to the interframe case. Statistical coding procedures

utilize the stochastic rule that exists among the elements that are to be coded.

The existence of strong nonstochastic dependence among elements of a multidimensional image implies that a purely statistical approach to image coding is suboptimal. The intraframe and color images can be sufficiently characterized by statistical means. The statistical approach can be extended to the interframe case, (as is done in the remaining subsections of this chapter). However, it is interesting to note various deterministic relations which apply to the interframe case and which are ignored by the statistical approach.

The above indicated deterministic rules can be formally represented by operator notation. Let $I_j(x, y) = I(x, y, t_j)$ and $I_k(x, y) = I(x, y, t_k)$ be two individual images in a sequence of images describing a time-dependent visual scene. Specifically, I_j and I_k represent the image at times t_j and t_k. The following specific question should be addressed: given the image pair I_j and I_k, is there a nonstochastic operator L, such that, at least approximately,

$$I_k(x, y) = L \left\{ I_j(x, y) \right\} \tag{6.1-1}$$

Any statistical coding approach which ignores Equation (6.1-1) and the inherent redundancy it implies cannot be optimal.

In the following, the various basic forms of the operator L are considered. The appropriate infiuence in the transform domain are addressed.

The operator L can represent local as well as global changes between I_j and I_k. The former case implies that only a relatively small part of the image is changing. Under global changes, the entire image is understood to be changing in some (nonstochastic) systematic manner. The local image variation indicates the temporal evolution of a visual scene as observed by a stationary imaging device. Global image variation is the probable result of the movement of the appropriate imaging device.

Some of the obvious global variations are image shift, rotation, defocus and magnification. Other more complicated global image changes as well as the simultaneous occurrence of the ones listed above clearly are possible. These global image variations cannot be characterized statistically; thus, the statistical encoder is not likely to remove the entire redundancy which is present in interframe imagery.

The extension of the adaptive phase coder of Chapter 4 to the interframe case is likely to be sub-optimal because of the statistical approach taken. For local variations and/or small global changes, the statistical correlation among neighboring frames is relatively high, thus utilization of the statistical approach will lead to modest bandwidth reduction over the intraframe approach.

Appropriate changes in the transform domain, resulting from the affects of the operator L in Equation (6.1-1), can be modeled by the use of simplified examples.

Local Variations

Consider the following case of frame-to-frame change as indicated in Figure 6.1-1. Let $g_3(x, y)$ represent a subregion in frame A which is shifted a distance $\underline{a}$ in the horizontal direction by the time that frame B is generated. The unchanging background is represented by $g_1(x, y)$. The altered parts of the background are denoted by $g_2(x, y)$ and $g_4(x, y)$. In frame A, $g_2(x, y)$ is part of the frame while $g_4(x, y)$ is covered by $g_3(x, y)$. The roles of $g_2(x, y)$ and $g_4(x, y)$ are interchanged in frame B. Equivalently, this can be expressed as

$$g_A = g_1 + g_2 + g_3 \qquad (6.1-2)$$

$$g_B = g_1 + g_3(x + a, y) + g_4 \qquad (6.1-3)$$

Here g_A and g_B represent frames A and B. Note also that the argument (x, y) is omitted for notational convenience. Although Equations (6.1-2) and (6.1-3) model a rather simplified interframe change, it approximates actual applications such as the Picturphone model. For the latter case, g_3 can be considered as the model for the Picturphone speaker and g_1, g_2 and g_4 represent the various background segments.

In terms of the previously developed notation, the frame-to-frame change can be made in a simpler form. First, the following additional definitions are made

$$g_a = g_2 + g_3 \qquad (6.1-4)$$
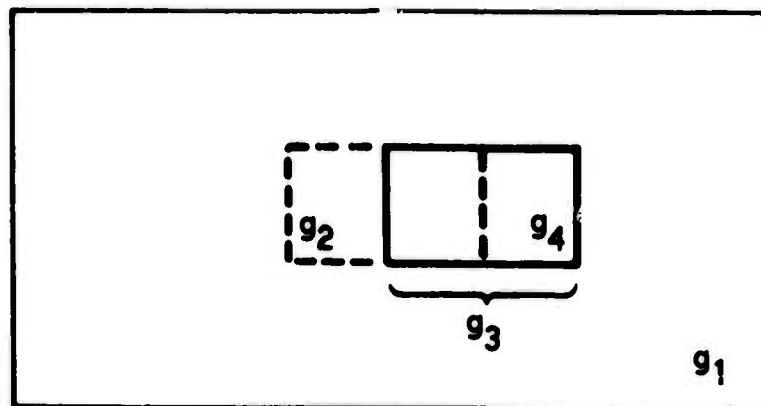
(a) Frame A
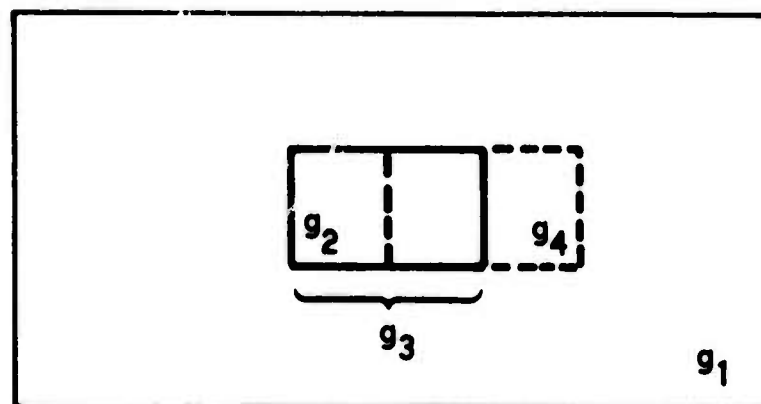
(b) Frame B

Figure 6.1-1.   Geometry of Frame Movements

$$g_b = g_3(x + a) + g_4 \qquad (6.1\text{-}5)$$

therefore

$$g_A = g_1 + g_a \qquad (6.1\text{-}6)$$

$$g_B = g_1 + g_b \qquad (6.1\text{-}7)$$

Equations (6.1-6) and (6.1-7) indicate that, for the model under discussion, each frame can be decomposed into a varying part and one that remains unaltered between consecutive frames.

Specifying the discussion to the Fourier transform, the above described model permits qualitative predictions for interframe changes in the frequency domain. Let $G_s(u, v)$ be the Fourier transform of $g_s(x, y)$. The subscript s may represent any of those previously utilized: 1, 2, 3, 4, a, b, A, and B. Consequently,

$$G_s(u, v) = G_s = \int_{-\infty}^{\infty} g_s(x, y) \exp - 2\pi j(ux + vy) \, dxdy$$

$$(6.1\text{-}8)$$

$$G_s = \left| G_s \right| \exp j\theta_s = G_{sR} + jG_{sI} \qquad (6.1\text{-}9)$$

$$\theta_s = \tan^{-1}\left(G_{sI}/G_{sR} \right) \qquad (6.1\text{-}10)$$

Note that the subscripts (u, v) are dropped whenever possible for notational convenience.

Using the developed notation for the frequency domain, the Fourier transforms of Equations (6.1-6) and (6.1-7) can be written as

$$G_A = G_1 + G_a \qquad (6.1-11)$$

$$G_B = G_1 + G_b \qquad (6.1-12)$$

The interframe variations in the frequency plane given by Equations (6.1-11) and (6.1-12) can be demonstrated by phasor diagrams. Simultaneous display of Equations (6.1-11) and (6.1-12) is shown in Figure 6.1-2a. This figure indicates a "typical" example of the interframe variation and should only be viewed as a qualitatively demonstrative example. The following assumptions are also inherent in this graphical demonstration: (1) g_1, g_a, and g_b have "similar" Fourier decompositions, (2) the region over which g_1 is defined is larger than the similarly-specified regions for g_a and g_b. The above assumptions imply that the power spectral density functions are similar for g_1, g_a and g_b except for different scale factors.

The graphical representation implies that both the amplitude and phase values are strongly correlated. Furthermore, the following inequalities for phase and amplitude changes are easily obtained from Figure 6.1-2a.
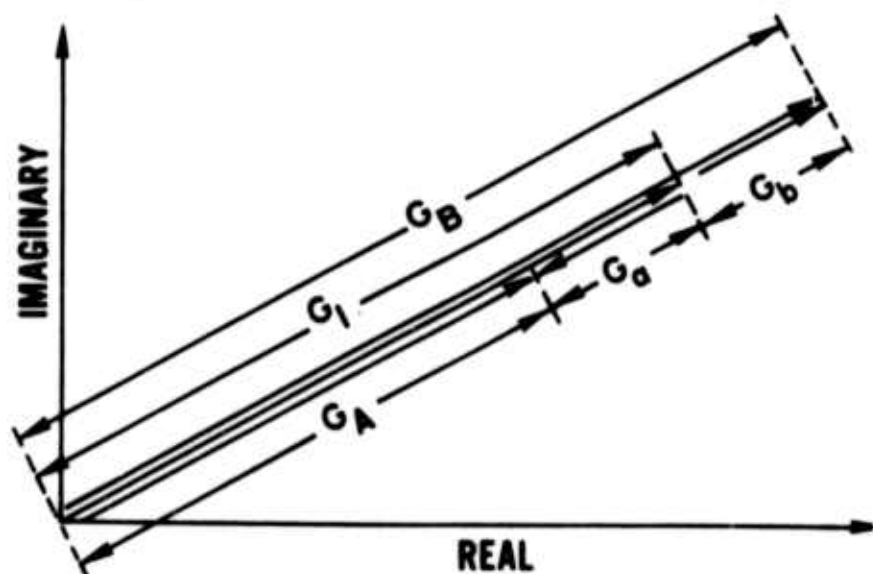
$$|\Delta\phi| = |\phi_A - \phi_B| \leq \left| \tan^{-1} \frac{G_b}{G_1} \right| + \left| \tan^{-1} \frac{G_a}{G_1} \right| \qquad (6.1-13)$$
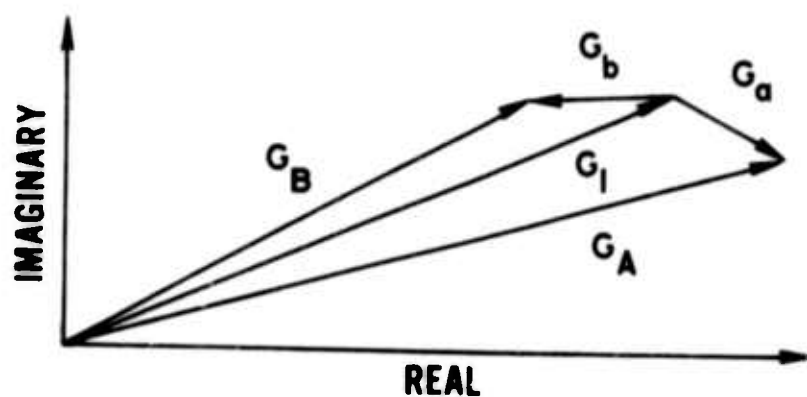
(a) Phasor Representation of Interframe Changes
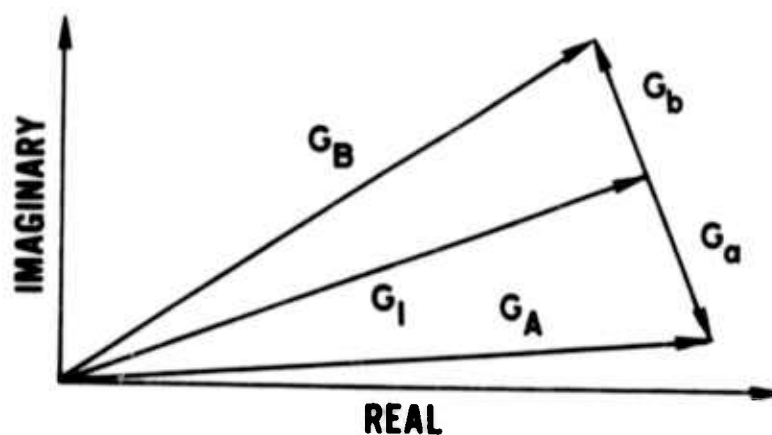
(b) Maximum Interframe Phase Deviation Between G_A and G_B

(c) Maximum Interframe Amplitude Deviation Between G_A and G_B

Figure 6.1-2. Vector Representation of Interframe Changes

$$|\Delta G| = \left| |G_A| - |G_B| \right| \leq |G_a| + |G_b| \qquad (6.1-14)$$

The conditions for equality in Equations (6.1-13) and (6.1-14) are illustrated in Figures 6.1-2b and c.

As a further demonstration of the phase and amplitude correlation, quantitative evaluation of Equations (6.1-13) and (6.1-14) can be made by substituting "reasonable" values for G_1, G_a, and G_b. In particular, let $|G_1| = |G_a| = |G_b|$. This condition implies that the areal extent of the changing and unchanging image segments are equal, with similar power spectral densities. Although the amplitude change constraint is not significant, as indicated by Equation (6.1-14), the maximum phase change is restricted to $\pi/2$ which is a fourfold reduction on the phase range.

For a 10 percent image area change, and similar assumptions as before, the application of Equation (6.1-14) indicates a phase change of less than $\pi/15$ which is a reduction of the phase range by a factor of 30.

6.2   Description of the Algorithm

The extension to the interframe case of the adaptive algorithm of Chapter 4 is structurally quite similar to the color coding implementation. The coding procedure utilizes the three-dimensional transform (Fourier or Walsh) to uncorrelate a set of four subsequent image frames. A schematic diagram similar to the one given for the color coder is shown in Figure 6.2-1.
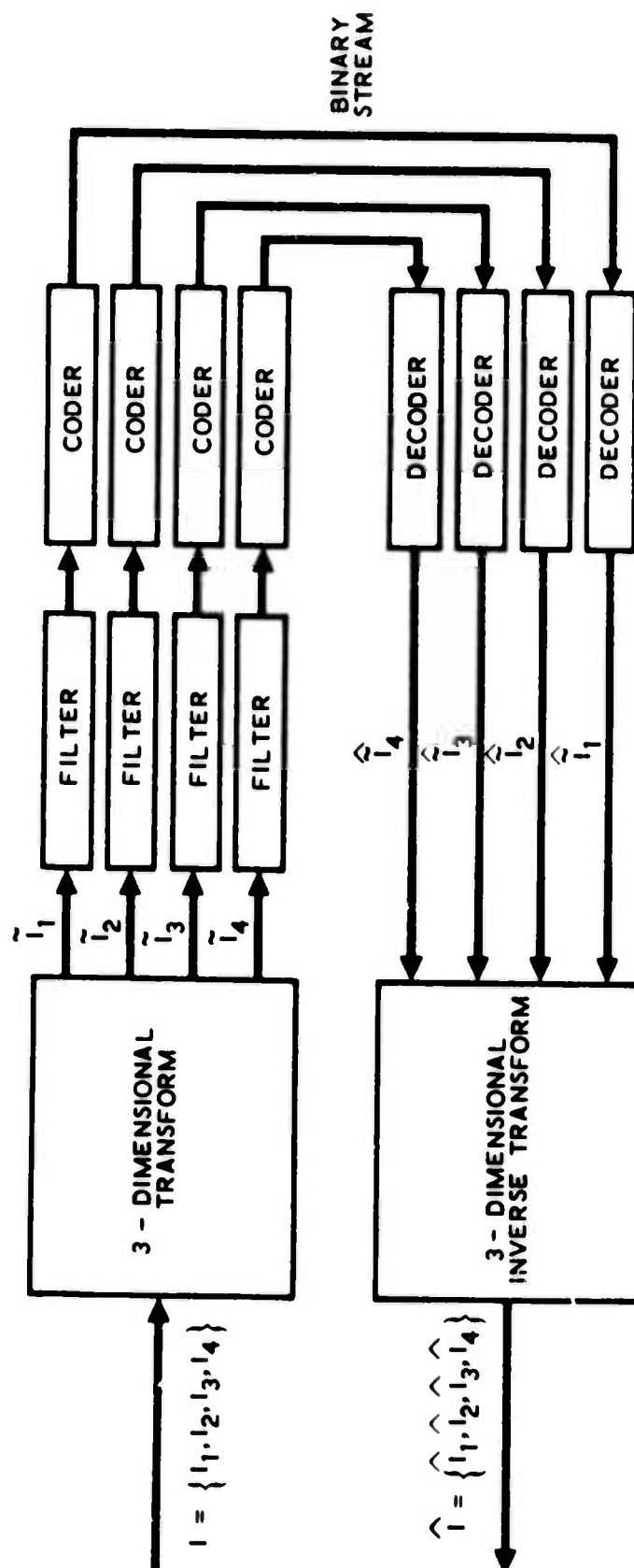
153

Figure 6.2-1. Schematic of Docing-Decoding Process for the Interframe Case

154

The three-dimensional transformation consists of the subsequent application of the two-dimensional 256 x 256 transform of each image plane and the one dimensional four-point transform along the temporal axis. The four-point transform matrices are shown in Figures 6.2-2 and 6.2-3 for the Fourier and Walsh matrices, respectively.

By assumption, the four transform planes are uncorrelated individually as well as relative to each other. After the application of the three-dimensional transform (either Fourier or Walsh) the first transform plane is the average of the two-dimensional transforms of the four input images. The other three transform planes represent fluctuations around the average. One could qualitatively argue that the first transform plane represents the unchanging image segment while the other three transform planes contain information relating to temporal variation.

Like the color coder of Chapter 5, each of the four transform planes is individually filtered and coded. Unlike, however, the color coding procedure, one cannot arbitrarily low-pass filter transform planes 2 through 4. Drastic low-pass filtering of these planes will result in the blurring of the time-varying areas without reducing the resolution of unchanging areas.

It has been demonstrated (Budrikas, 1972) that the resolution loss in rapidly changing areas is visually much less objectionable than for image segments that are relatively stationary. By

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & -1 & -j \\ 1 & -1 & 1 & -1 \\ 1 & -j & -1 & j \end{bmatrix}$$

Figure 6.2-2. Four-Element Fourier Transform

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

Figure 6.2-3. Four-Element Walsh Transform

appropriately "tuning" the four two-dimensional filters in Figure 6.2-2, the psychophysical properties of the human visual system could be exploited. Although the computer-implemented algorithm of this chapter could be utilized for the study of the relative importance of resolution loss in moving and stationary image segments, it was not done experimentally. The unavailability to this research effort of the hardware required to display the decoded interframe images in their natural medium (such as television) restricted the visual evaluation of the decoded image sequences to the viewing of individual (stationary) images.

The structure of both the Fourier and Walsh transform matrices indicates that for statistically correlated image frames the image energy will concentrate in the first transform plane. Therefore, even without the application of different spatial filters to the various transform planes, the adaptive procedure will result in bandwidth reduction. The transform values in transform planes 2 through 4 will require fewer quantization levels because of uneven energy distribution.

The advantages of the adaptive phase coding procedure indicated in Chapters 4 and 5, are applicable to the interframe coder as well. Specifically, the coder will "track" the three-dimensional power spectrum and make the bit assignment adaptively. The number and location of quantum levels will be specified according to the local estimated value of the power spectral density. The adaptivity feature

has an added benefit for the interframe transform coder. Unlike the monochrome exponential correlation model, the interframe case cannot be modeled by a simple correlation function. In fact, the highly non-stationary nature of the interframe image precludes any fixed nonadaptive modeling of the transform domain.

6.3 Pictorial Examples

Examples for the three-dimensional Fourier and Walsh interframe coder are shown in this subsection. Figure 6.3-1 is the three-dimensional transform domain display. The coding examples are given in Figures 6.3-2 through 6.3-9. An example for the decoded transform planes is given in Figure 6.3-10.

The visual inspection of Figures 6.3-1 and 6.3-10 demonstrate the "non-stationary" character of the three-dimensional transform for the interframe case and the capability of the coding method to adapt to the particular form. The structure of transform planes 2 through 4 is the result of the significant amount of image motion in this example.
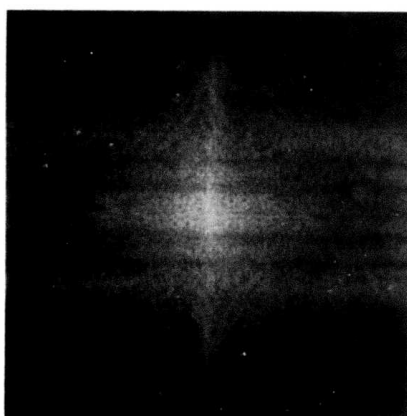
The Fourier transform coder similarly to the monochrome case outperforms the Walsh coder both in terms of mean square error as well as visual appearance.
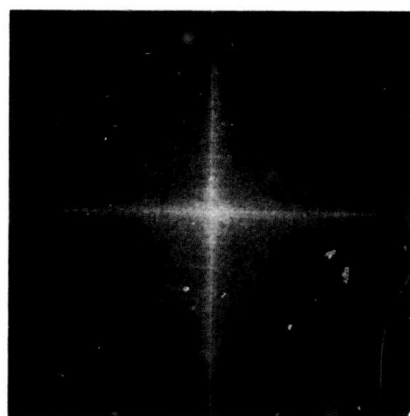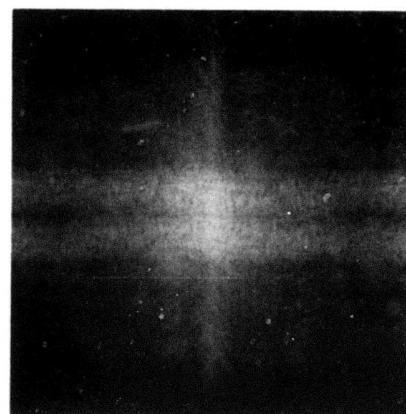
(a) First Plane
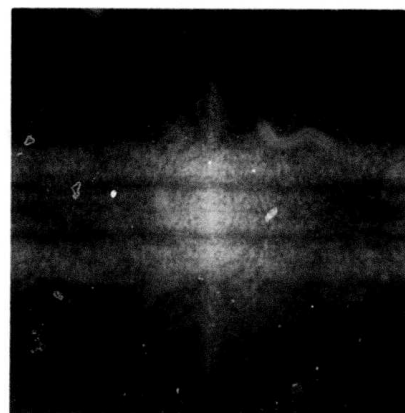
(b) Second Plane

(c) Third Plane

(d) Fourth Plane

Figure 6.3-1. Three-Dimensional Fourier Transform Display
of BELL DUMMY Image Sequence

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-2. Decoded BELL-GIRL Image Sequence I (Fourier).
Bit Rate: 0.27 Bit; M.S.E.: 1.79% (normalized
relative to image energy)

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-3. Decoded BELL-GIRL Image Sequence, II (Fourier).
Bit Rate 0.55 Bit; M.S.E.: 0.99% (normalized
relative to image energy)

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-4.   Decoded BELL-GIRL Image Sequence,
III (Walsh).   Bit Rate 0.38 Bit; M.S.E.:   2.24%
(Normalized Relative to Image Energy)

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-5. Decoded BELL-GIRL Image Sequence,
IV (Walsh). Bit Rate 0.68 Bit, M.S.E.: 2.15%
(Normalized Relative to Image Energy)

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-6.   Decoded BELL-DUMMY Image Sequence,
I (Fourier).   Bit Rate: 0.26 Bit;
M.S.E.: 1.47% (normalized
relative to image energy)

This page is reproduced at the
back of the report by a different
reproduction method to provide
better detail.

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-7. Decoded BELL-DUMMY Image Sequence,
II (Fourier). Bit Rate: 0.43 Bit; M.S.E.: 1.05%
(normalized relative to image energy)

(a)   First Image   (b)   Second Image

(c)   Third Image   (d)   Fourth Image

Figure 6.3-8.   Decoded BELL-DUMMY Image Sequence,
III (Walsh).   Bit Rate: 0.29 Bit; M.S.E.: 2.24%
(normalized relative to image energy)
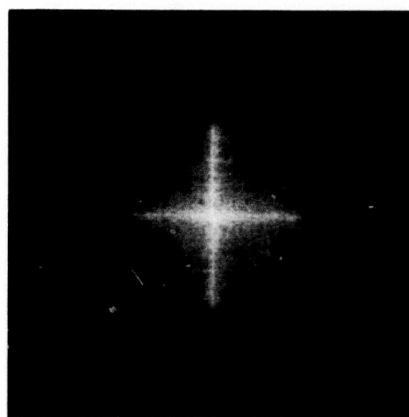
(a) First Image
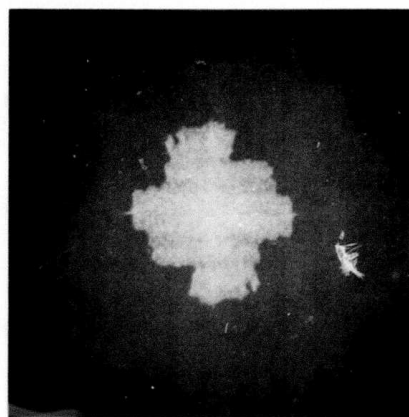
(b) Second Image

(c) Third Image

(d) Fourth Image

Figure 6.3-9. Decoded BELL-DUMMY Image Sequence,
IV (Walsh). Bit Rate 0.69 Bit; M.S.E.: 1.69%
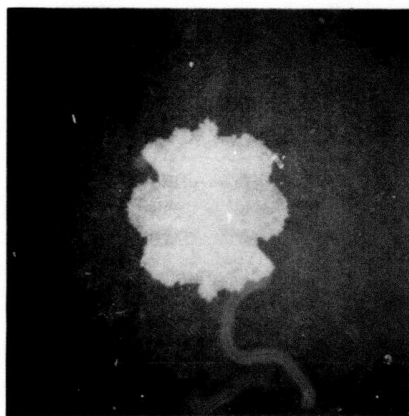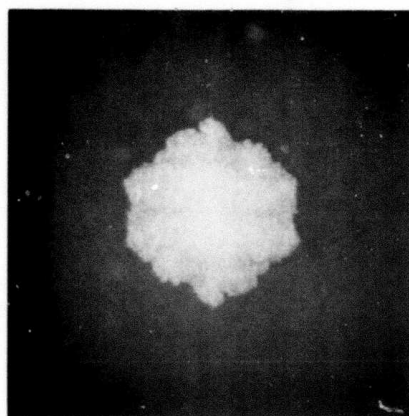(normalized relative to image energy)

(a) First Plane

(b) Second Plane

(c) Third Plane

(d) Fourth Plane

Figure 6.3-10.   Decoded Transform Planes for BELL DUMMY
Image Sequence

7. SUMMARY

A new approach to transform image coding has been presented in this dissertation. The generalized phase concept plays a dominant role in currently developed coding algorithms. The important advantage of the coding algorithms is a high degree of adaptivity in the determination of both the number and location of the quantum levels.

Significant adaptivity to the image power spectral density is accomplished without the assumption or specification of any a priori statistical image model. In addition, no bookkeeping information is required. The actual image model is "dynamically" determined from previously quantized and reconstructed transform samples.

The new transform coding approach was implemented through discrete Walsh and Fourier transforms. The Fourier transform was found to be superior to the Walsh transform. The fundamental superiority of the Fourier transform is explained by the general image insensitivity to (frequency domain) low-pass filtering.

Although the image transform is performed on the entire (256×256) image rather than on smaller blocks, the increase in computational complexity is modest. For example, the number of computational steps will only increase a factor of two from 16×16 block transforms to the entire 256×256 transform. The large-size transform is a disadvantage if a hard-wired configuration is required, however, this fact is unimportant when the coding-decoding algorithm is implemented via general purpose computers. This latter case has

168

the potential practical utility for computer-to-computer image transmission.

The new image coding techniques utilizing large-size transforms significantly outperform the block-encoding transform technique. The usual block size (16 X 16) exceeds the number of picture elements over which the image is significantly correlated; it was therefore previously postulated that larger-size transform blocks may result in negligible performance improvement. However, block encoding, particularly at low data rates, assigns a significant fraction of the available bits for reconstruction of block-to-block boundaries. Stated in another way, the image statistics are significantly altered by grouping into adjacent image blocks. Discovery and analysis of this fact provides a sound theoretical basis for the experimental success of the coding algorithms in this dissertation.

The experimental portion of the dissertation includes coding algorithms for monochrome, color, and interframe images. It has been found that the data rate can decrease to 0.38 bit for monochrome, 0.55 bit for color and 0.25 bit for interframe images. The implementation included both the Fourier and Walsh transforms. Visual image degradation, however, was more significant for the Walsh than for the Fourier transform.

The coding scheme is susceptible to channel errors. It was shown that the coder output is statistically equivalent to a discrete memoryless source, thus, conventional channel encoding techniques are applicable. The coding procedure is capable for a wide range of

data compression thus the requirement for algebraic redundancy (channel coding) can be offset by additional image data compression (source coding).

The non-negative image constraint has been studied via the Lukosz bound.

The conclusions of this dissertation are

a) Determination of the proper transform domain image model is important.

b) Utilization of large-size transforms and adaptive phase coding permits significant additional rate reduction when comparison is made with block encoding.

c) The superiority of phase has been demonstrated as a random variable for coding.

d) The development of improved predicting algorithms and preprocessing filters may result in additional bandwidth reduction. The polynomial surface fit algorithm, in addition, could be utilized for the image model.

e) Adaptivity is important to deal with non-stationary image structure, particularly for the interframe case, and residual noise. The latter consideration was shown to be important for most practical situations.

APPENDIX A

ORIGINAL TEST IMAGES

Various test images used in experimental sections of this
dissertation are shown in this appendix. Figure A-1 shows redisplay
of the two original monochrome images. Color test images are
shown in Figure A-2. Their three primary components (tristimulus
values in the NTSC receiver phosphor primary system) are shown in
Figure A-3 in monochrome presentations. Two image sequences
used for interframe coding are shown in Figures A-4 and A-5.
Image differences for these sequences are presented in Figure A-6.

The monochrome (Figure A-1) and color test images (Figure
A-2) were obtained by digitization of photographic transparencies.
The image sequences for the interframe case were obtained from
digitized video signal. All sampled images consist of 256 × 256
picture elements and each original sample is uniformly quantized to
256 levels (8 bits). The monochrome images were displayed on a
flying spot scanner and photographed on Polaroid-type 52 film. The
color images were displayed on the Aerojet Model SG-D2219 video
display and photographed on high-speed Ektachrome film.

(a) Girl

(b) Couple

This page is reproduced at the
back of the report by a different
reproduction method to provide
better detail.

Figure A-1. Monochrome Test Images

(a) Girl

(b) Couple

Figure A-2. Color Test Images

This page is reproduced at the
back of the report by a different
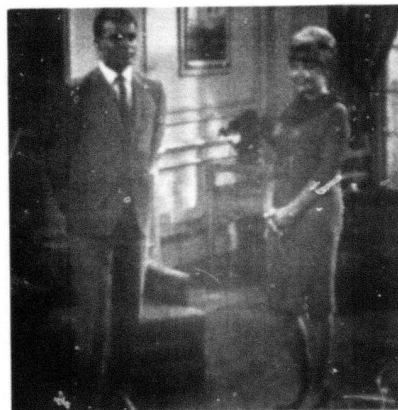reproduction method to provide
better detail.

Figure A-3. Primary Component Images. (a) Red, (b) Green, (c) Blue

This page is reproduced at the back of the report by a different reproduction method to provide better detail.

(a) First Image

(b) Second Image

(c) Third Image

(d) Fou:th Image

Figure A-4.   Image Sequence: BELL-GIRL

(a) First Image

(b) Second Image

(c) Third Image

(d) Fourth Image

Figure A-5. Image Sequence: BELL DUMMY

(a) Difference Between First and Second Image

(b) Difference Between Second and Third Image

(c) Difference Between Third and Fourth Image

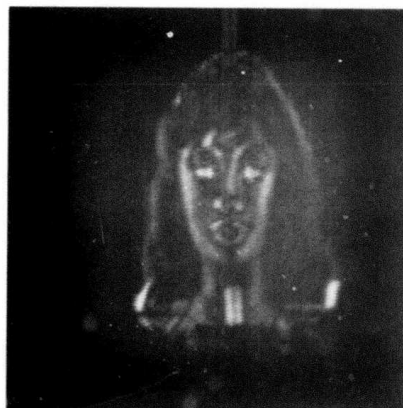Figure A-6. Absolute Image Differences Among Consecutive
Images in Image Sequences of
Figures A-4 and A-5

This page is reproduced at the
back of the report by a different
reproduction method to provide
better detail.

APPENDIX B

NUMERICAL NOISE GENERATED BY LARGE-
SIZE FOURIER TRANSFORMS

Results of a simple computer experiment are shown in this appendix, to demonstrate that large (256 × 256) numerical Fourier transforms are expected to generate a negligible amount of numerical noise. The "girl" image was Fourier-transformed and then the result inverse-transformed and the appropriate mean squared error was calculated. This cycle was repeated two more times on the retransformed images. The results are shown in Table B-1. All calculations were performed on an IBM 360/44 computer with single precision (32 bit) arithmetic.

TABLE B-1

DEMONSTRATION OF FOURIER TRANSFORM-
GENERATED NUMERICAL NOISE

Cycle	MSE
1	$9 \times 10^{-8}\%$
2	$49 \times 10^{-8}\%$
3	$100 \times 10^{-8}\%$

Table B-1 demonstrates that the large-size transform-generated generated numerical noise probably will have negligible influence on image coding problems.

REFERENCES

Anderson, G. B., and Huang, T. S., 1971, "Piecewise Fourier Transformation for Picture Bandwidth Compression." IEEE Trans. Communication Technology. COM-19, pp. 133-140

Andrews, H. C., 1970, Computer Techniques in Image Processing. Academic Press.

Bhushan, A. K., 1970, "Transmission and Coding of Color Pictures." in Picture Bandwidth Compression. eds. T. S. Huang and O. J. Tretiak. Gordon and Breach.

Bingham, G., Godfrey, M. D., and Tukey, J. W., 1967, "Modern Techniques of Power Spectrum Estimation." IEEE Trans. Audio and Electroacoustics. Au-15, pp. 56-66.

Blackman, R. B. and Tukey, J. W., 1959, The Measurement of Power Spectra. Dover.

Blum, E. K., 1972, Numerical Analysis and Computation: Theory and Practice. Addison-Wesley Publishing Company.

Budrikas, Z. L., 1972, "Visual Fidelity Criterion and Modeling." Proc. IEEE. 70, pp. 771-779.

Chu, D. C., and Goodman, J. W., 1972, "Spectrum Shaping with Parity Sequences." Applied Optics, 11, pp. 1716-1724.

Chu, D. C., Fienup, J. R., and Goodman, J. W., 1973, "Multi-emulsion On-Axis Computer Generated Hologram." Applied Optics, 12, 1386-1388.

Cooley, J. W., and Tukey, J. W., 1965, "An Algorithm for the Machine Calculation of Fourier Series." Math. of Complet., 19, pp. 297-301.

Cornsweet, T. N., 1970, Visual Perception. Academic Press.

Courant, R., and Hilbert, D., 1953, Methods of Mathematical Physics. Interscience.

Cutler, C. C., 1952, Patent No. 2,605,361, July 29, 1952.

Dallas, W. J., 1971 (a), "Phase Quantization - A Compact Derivation." Applied Optics, 10, pp. 673-674.

Dallas, W. J., 1971 (b), "Phase Quantization in Holograms - A Few Illustrations." Applied Optics, 10, pp. 674-675.

Davenport, W. B., and Root, W. L., 1958, An Introduction to the Theory of Random Signals and Noise. McGraw-Hill Book Company, Inc.

Franks, L. E., 1966, "A Model for Random Video Process." Bell Syst. Tech. J. 45, pp. 609-630.

Frei, A. H., Schindler, H. R., and Vectinger, P., 1972, "An Adaptive Dual-Mode Coder/Decoder for Television Signals." IEEE Trans. on Communication Technology, Com-19, pp. 933-944.

Gallager, R. G., 1968, Information Theory and Reliable Communication. John Wiley and Sons, Inc.

Goodman, J. W., 1968, Introduction to Fourier Optics. McGraw-Hill Book Co.

Goodman, J. W. and Silvesteri, A. M., 1970, "Some Effects of Fourier-Domain Phase Quantization." IBM J. of Res. & Develop. pp. 478-484.

Graham, R. E., 1958, "Predictive Quantizing of Television Signals." IRE Wescom Conv. Rec., Part 4, pp. 147-157.

Graham, D. N., 1968, "Image Transmission by Two-Dimensional Contour Coding." Proc. IEEE, 55, pp. 336-346.

Habibi, A., and Wintz, P. A., 1971, "Image Coding by Linear Transformations and Block Quantization." IEEE Trans. Comm. Tech., COM-19, pp. 50-60.

Habibi, A., 1971, "Comparison of n-th Order DPCM Encoder With Linear Transformation and Block Quantization Techniques." IEEE Trans. on Comm. Tech., COM-19, pp. 50-63.

Habibi, A., 1972, "Quantization Error and Entropy of a Single Random Variate." Semiannual Technical Report USCEE 425.

Habibi, A., 1973, "Coding Television by Contour Tracing the Differential Signals." Semiannual Technical Report USCEE 444.

Harmuth, H. F., 1972, Transmission of Information by Orthogonal Functions. Springer-Verlag.

Haskell, B. G., Mounts, F. W., and Candy, J. C., "Interface Coding of Videotelephone Pictures." Proc. IEEE, 60, pp. 792-800.

Hoffman, K., and Kunze, R., 1961, Linear Algebra. Prentice-Hall.

Huang, T. S., 1971, "Digital Holography." Proc. IEEE, 59, 1335-1346.

Huang, T. S., Schreiber, W. F., and Tretiak, O. J., 1971, "Image Processing." Proc. IEEE, 59, pp. 1586-1609 (1971).

Kermisch, D., 1970, "Image Reconstruction from Phase Information Only." J. Opt. Soc. Am., 60, pp. 15-17.

Levi, L., 1970, "On Image Evaluation and Enhancement." Optica Octa, 17, pp. 59-76.

Lukosz, W., 1962, "Ubertragung Nicht-Negativer Signale durch Lineare Filter." Optica Octa, 9, pp. 335-364.

Max, J., 1960, "Quantizing for Minimum Distortion." IRE Trans. Inf. Theory, IT-6, pp. 7-12.

O'Neill, E. L., 1963, Introduction to Statistical Optics. Addison-Wesley Publishing Co.

O'Neal Jr., J. B., 1966, "Predictive Quantizing Systems for the Transmission of Television Signals." BSTJ 45, pp. 689-722.

Oppenheim, A. V., Schafer, R. W., and Stockham, T. G., 1968, "Nonlinear Filtering of Multiplied and Convolved Signals." Proc. IEEE, 56, pp. 1264-1291.

Papoulis, A., 1965, Probability, Random Variables, and Stochastic Processes. McGraw-Hill Book Company.

Panter, P. F., and Dite, W., 1951, "Quantization Distortion in Pulse-Code Modulation with Nonuniform Spacing of Levels." Proc. IRE 39, pp. 44-48.

Pratt, W. K., 1971, "Spatial Transform of Color Images." IEEE Trans. Comm. Tech., COM-19, pp. 980-991 (1971).

Pratt, W. K., 1972, "Generalized Wiener Filtering Computation Techniques." IEEE Trans. Comp., C-21, pp. 636-641.

Pratt, W. K., Welch, L. R., and Chen, W., 1972, "Slant Transforms for Image Coding." Proc. Symposium on Applications of Walsh Functions. Washington, D.C., pp. 229-234.

Rosenfeld, A., 1969, Picture Processing by Computer, Academic Press.

Shannon, C. E., and Weaver, W., 1949, The Mathematical Theory of Communication. Univ. of Illinois Press.

Smith, B., 1957, Instantaneous Companding of Quantized Signals. Bell Sys. Tech. J. 36, pp. 653-709.

Stockham, Jr., T. G., 1972, "Image Processing in the Context of a Visual Model." Proc. IEEE, 60, pp. 828-842.

Thomas, J. B., 1968, An Introduction to Statistical Communication Theory. John Wiley and Sons, Inc.